Notes on a PDE System for Biological Network Formation

Jan Haskovec, Peter Markowich, Benoît Perthame, Matthias Schlottbom

Notes on a PDE System for Biological Network Formation

Jan Haskovec[1] Peter Markowich[2] Benoît Perthame[3] Matthias Schlottbom[4]

Abstract. We present new analytical and numerical results for the elliptic-parabolic system of partial differential equations proposed by Hu and Cai [8, 10], which models the formation of biological transport networks. The model describes the pressure field using a Darcy's type equation and the dynamics of the conductance network under pressure force effects. Randomness in the material structure is represented by a linear diffusion term and conductance relaxation by an algebraic decay term. The analytical part extends the results of [7] regarding the existence of weak and mild solutions to the whole range of meaningful relaxation exponents. Moreover, we prove finite time extinction or break-down of solutions in the spatially one-dimensional setting for certain ranges of the relaxation exponent. We also construct stationary solutions for the case of vanishing diffusion and critical value of the relaxation exponent, using a variational formulation and a penalty method.

The analytical part is complemented by extensive numerical examples. We propose a discretization based on mixed finite elements and study the qualitative properties of network structures for various parameters values. Furthermore, we indicate numerically that some analytical results proved for the spatially one-dimensional setting are likely to be valid also in several space dimensions.

Key words: Network formation; Weak solutions; Stability; Penalty method; Numerical experiments
Math. Class. No.: 35K55; 35B32; 92C42

Contents

[1]Mathematical and Computer Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia; *jan.haskovec@kaust.edu.sa*

[2]Mathematical and Computer Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia; *peter.markowich@kaust.edu.sa*

[3]Sorbonne Universits, UPMC Univ Paris 06, Inria, Laboratoire Jacques-Louis Lions UMR CNRS 7598, F-75005, Paris, France, *benoit.perthame@upmc.fr*

[4]Institute for Computational and Applied Mathematics, University of Münster, Einsteinstr. 62, 48149 Münster, Germany; *schlottbom@uni-muenster.de*

1 Introduction

In [7] we presented a mathematical analysis of the PDE system modeling formation of biological transportation networks

$$-\nabla \cdot [(rI + m \otimes m)\nabla p] = S, \tag{1.1}$$

$$\frac{\partial m}{\partial t} - D^2 \Delta m - c^2 (m \cdot \nabla p)\nabla p + \alpha |m|^{2(\gamma-1)} m = 0, \tag{1.2}$$

for the scalar pressure $p = p(t,x) \in \mathbb{R}$ of the fluid transported within the network and vector-valued conductance $m = m(t,x) \in \mathbb{R}^d$ with $d \leq 3$ the space dimension. The parameters are $D \geq 0$ (diffusivity), $c > 0$ (activation parameter), $\alpha > 0$ and $\gamma \in \mathbb{R}$; in particular, we restricted ourselves to $\gamma \geq 1$ in [7]. The scalar function $r = r(x) \geq r_0 > 0$ describes the isotropic background permeability of the medium. The source term $S = S(x)$ is assumed to be independent of time and $\gamma \in \mathbb{R}$ is a parameter crucial for the type of networks formed [10]. In particular, experimental studies of scaling relations of conductances (diameters) of parent and daughter edges in realistic network modeling examples suggest that $\gamma = 1/2$ can be used to model blood vessel systems in the human body and $\gamma = 1$ is adapted to leaf venation [8, 9]. For the details on the modeling which leads to (1.1), (1.2) we refer to [1].

The system was originally derived in [8, 10] as the formal gradient flow of the continuous version of an energy functional describing formation of biological transportation networks on discrete graphs. We pose (1.1), (1.2) on a bounded domain $\Omega \subset \mathbb{R}^d$ with smooth boundary $\partial\Omega$, subject to homogeneous Dirichlet boundary conditions on $\partial\Omega$ for m and p:

$$m(t,x) = 0, \qquad p(t,x) = 0 \qquad \text{for } x \in \partial\Omega,\ t \geq 0, \tag{1.3}$$

and subject to the initial condition for m:

$$m(t=0,x) = m^0(x) \qquad \text{for } x \in \Omega. \tag{1.4}$$

The main mathematical interest of the PDE system for network formation stems from the highly unusual nonlocal coupling of the elliptic equation (1.1) for the pressure p to the reaction-diffusion equation

(1.2) for the conductance vector m via the pumping term $+c^2(\nabla p \otimes \nabla p)m$ and the latter term's potential equilibriation with the decay term $-|m|^{2(\gamma-1)}m$. A major observation concerning system (1.1)–(1.2) is that it represents the formal $L^2(\Omega)$-gradient flow associated with the highly non-convex energy-type functional

$$\mathcal{E}(m) := \frac{1}{2}\int_\Omega \left(D^2|\nabla m|^2 + \frac{\alpha}{\gamma}|m|^{2\gamma} + c^2|m\cdot\nabla p[m]|^2 + c^2 r(x)|\nabla p[m]|^2 \right) \mathrm{d}x, \tag{1.5}$$

where $p = p[m] \in H_0^1(\Omega)$ is the unique solution of the Poisson equation (1.1) with given m, subject to the homogeneous Dirichlet boundary condition on $\partial\Omega$. We have:

Lemma 1 (Lemma 1 in [7])**.** *Let $\mathcal{E}(m^0) < \infty$. Then the energy $\mathcal{E}(m(t))$ is nonincreasing along smooth solutions of* (1.1)–(1.2) *and satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(m(t)) = -\int_\Omega \left(\frac{\partial m}{\partial t}(t,x)\right)^2 \mathrm{d}x.$$

As usual, along weak solutions, we obtain a weaker form of energy dissipation, see formula (2.4) below. In [7] we provided the following analytical results for (1.1)–(1.4) in the case $\gamma \geq 1$:

- Existence of global weak solutions in the energy space
- Existence and uniqueness of local in time mild solutions (global in 1d)
- Existence of nontrivial (i.e., $m \not\equiv 0$) stationary states and analysis of their stability (nonlinear in 1d, linearized in multiple dimensions)
- The limit $D \to 0$ in the 1d setting

The purpose of this paper is to extend the analysis of the network formation system by providing several new results, in particular:

- Existence of global weak solutions in the energy space for $1/2 \leq \gamma < 1$ and of local in time mild solutions for $1/2 < \gamma < 1$ (Section 2).
- Analysis of the system in the 1d setting: finite time breakdown of solutions for $\gamma < 1/2$, infinite time extinction for $1/2 \leq \gamma \leq 1$ with small sources, nonlinear stability analysis for $\gamma \geq 1/2$ and $D = 0$ (Section 3).
- Construction of stationary solutions in the case $\gamma = 1$ and $D = 0$ (Section 4).

The analytical part is complemented by extensive numerical examples in Section 5. We propose a discretization based on mixed finite elements and study the qualitative properties of network structures for various parameters values. Furthermore, we indicate numerically that some analytical results proved for the spatially one-dimensional setting are likely to be valid also in several space dimensions.

1.1 Scaling analysis

We introduce the rescaled variables

$$x_s := \frac{x}{\overline{x}}, \qquad t_s := \frac{t}{\overline{t}}, \qquad m_s := \frac{m}{\overline{m}}, \qquad p_s := \frac{p}{\overline{p}}, \qquad S_s := \frac{S}{\overline{S}}$$

and choose

$$\overline{x} := \mathrm{diam}(\Omega), \qquad \overline{m} := \sup_{x\in\Omega} |m^0(x)|, \qquad \overline{t} := \frac{1}{\alpha \overline{m}^{2(\gamma-1)}}, \qquad \overline{S} := \sup_{x\in\Omega} |S(x)|, \qquad \overline{p} := \frac{\overline{x}^2 \overline{S}}{\overline{m}^2}$$

which leads to $S_s = \mathcal{O}(1)$, $m_s(t=0) = \mathcal{O}(1)$ and the following rescaled version of (1.1)–(1.2),

$$\begin{aligned} -\nabla_{x_s} \cdot [(r_s I + m \otimes m) \nabla_{x_s} p_s] &= S_s, \\ \frac{\partial m_s}{\partial t} - D_s^2 \Delta_{x_s} m_s - c_s^2 (m_s \cdot \nabla_{x_s} p_s) \nabla p_s + |m_s|^{2(\gamma-1)} m_s &= 0, \end{aligned}$$

with

$$r_s = \frac{r}{\overline{m}^2}, \qquad D_s^2 = \frac{\overline{p}\,\overline{m}^2}{\overline{x}^2 \overline{s}}, \qquad c_s^2 = \frac{c^2 \overline{p}^2}{\alpha \overline{x}^2 \overline{m}^{2(\gamma-1)}}.$$

Dropping the index s in the scaled variables, we obtain the system

$$-\nabla \cdot [(rI + m \otimes m) \nabla p] = S, \tag{1.6}$$

$$\frac{\partial m}{\partial t} - D^2 \Delta m - c^2 (m \cdot \nabla p) \nabla p + |m|^{2(\gamma-1)} m = 0, \tag{1.7}$$

that we will study in this paper. Moreover, for simplicity, we set $r(x) \equiv 1$ in the analytical part (Sections 2–4).

Convention. *In the following, generic, not necessarily equal, constants will be denoted by C. Moreover, we will make specific use of the Poincaré constant C_Ω, i.e.,*

$$\|u\|_{L^2(\Omega)} \le C_\Omega \|\nabla u\|_{L^2(\Omega)} \qquad \textit{for all } u \in H_0^1(\Omega).$$

2 Existence of global weak solutions for $1/2 \le \gamma < 1$

In [7], Section 2, we provided the proof of existence of global weak solutions for $\gamma \ge 1$ based on the Leray-Schauder fixed point theorem for a regularized version of (1.6)–(1.7) that preserves the energy dissipation structure, and consequent limit passage to remove the regularization. We now extend the proof to the case $1/2 \le \gamma < 1$. However, the case $\gamma = 1/2$ requires special care since the algebraic term in (1.7) formally becomes $m/|m|$ and an interpretation has to be given for $m = 0$. In particular, (1.7) has to be substituted by the differential inclusion

$$\partial_t m - D^2 \Delta m - c^2 (m \cdot \nabla p[m]) \nabla p[m] \in -\partial \mathcal{R}(m), \tag{2.1}$$

where $\partial\mathcal{R}$ is the subdifferential of $\mathcal{R}(m) := \int_\Omega |m| \,\mathrm{d}x$, in particular,

$$\partial \mathcal{R}(m) = \{u \in L^\infty(\Omega)^d;\ u(x) = m(x)/|m(x)| \text{ if } m(x) \ne 0, \tag{2.2}$$

$$|u(x)| \le 1 \text{ if } m(x) = 0\}. \tag{2.3}$$

Theorem 1 (Extension of Theorem 1 of [7]). *Let $\gamma \geq 1/2$, $S \in L^2(\Omega)$ and $m^0 \in H^1_0(\Omega)^d \cap L^{2\gamma}(\Omega)^d$. Then the problem (1.6)–(1.7), (1.3)–(1.4) (with (2.1) instead of (1.7) if $\gamma = 1/2$) admits a global weak solution $(m, p[m])$ with $\mathcal{E}(m) \in L^\infty(0,\infty)$ and with*

$$\begin{aligned} m \in L^\infty(0,\infty; H^1_0(\Omega)) \cap L^\infty(0,\infty; L^{2\gamma}(\Omega)), \qquad & \partial_t m \in L^2((0,\infty)\times\Omega), \\ \nabla p \in L^\infty(0,\infty; L^2(\Omega)), \qquad & m \cdot \nabla p \in L^\infty(0,\infty; L^2(\Omega)). \end{aligned}$$

This solution satisfies the energy dissipation inequality, with $\mathcal{E}$ given by (1.5),

$$\mathcal{E}(m(t)) + \int_0^t \int_\Omega \left(\frac{\partial m}{\partial t}(s,x)\right)^2 \mathrm{d}x\,\mathrm{d}s \leq \mathcal{E}(m^0) \qquad \text{for all } t \geq 0. \tag{2.4}$$

The proof proceeds along the lines of Section 2 of [7], i.e., for $\gamma > 1/2$ and $\varepsilon > 0$ we consider the regularized system

$$-\nabla \cdot [\nabla p + m(m\cdot\nabla p) * \eta_\varepsilon] = S, \tag{2.5}$$

$$\frac{\partial m}{\partial t} - D^2\Delta m - c^2[(m\cdot\nabla p) * \eta_\varepsilon]\nabla p + |m|^{2(\gamma-1)}m = 0, \tag{2.6}$$

with $(\eta_\varepsilon)_{\varepsilon>0}$ the d-dimensional heat kernel $\eta_\varepsilon(x) = (4\pi\varepsilon)^{-d/2}\exp(-|x|^2/4\varepsilon)$ and m, p are extended by 0 outside Ω so that the convolution is well defined. For $\gamma = 1/2$, equation (2.6) has to be substituted by the differential inclusion

$$\frac{\partial m}{\partial t} - D^2\Delta m - c^2[(m\cdot\nabla p) * \eta_\varepsilon]\nabla p \in -\partial\mathcal{R}(m).$$

Weak solutions of the regularized system are constructed by an application of the Leray-Schauder fixed point theorem as in Section 2 in [7], the only change that needs to be done is a slight modification of the proof of Lemma 3 in [7]. In particular, for $\gamma > 1/2$ we construct weak solutions of the auxiliary problem

$$\frac{\partial m}{\partial t} - D^2\Delta m - f = -|m|^{2(\gamma-1)}m, \tag{2.7}$$

subject to the initial and boundary conditions

$$m(t=0) = m^0 \quad \text{in } \Omega, \qquad m = 0 \quad \text{on } \partial\Omega. \tag{2.8}$$

Again, for $\gamma = 1/2$ we have to consider the following differential inclusion instead,

$$\frac{\partial m}{\partial t} - D^2\Delta m - f \in -\partial\mathcal{R}(m). \tag{2.9}$$

In the subsequent lemma we construct the so-called *slow solution* of (2.9), which is the unique weak solution of the PDE

$$\frac{\partial m}{\partial t} - D^2\Delta m - f = -r(m) \tag{2.10}$$

with

$$[r(m)](x) = \begin{cases} m(x)/|m(x)| & \text{when } m(x) \neq 0, \\ 0 & \text{when } m(x) = 0. \end{cases} \tag{2.11}$$

Lemma 2 (Extension of Lemma 3 in [7]). *For every $D > 0$, $\gamma > 1/2$, $T > 0$ and $f \in L^2((0,T)\times\Omega)^d$, the problem (2.7)–(2.8) with $m^0 \in H_0^1(\Omega)^d$ admits a unique weak solution $m \in L^\infty(0,T;H_0^1(\Omega))^d \cap L^2(0,T;H^2(\Omega))^d \cap L^\infty(0,T;L^{2\gamma}(\Omega))^d$ with $\partial_t m \in L^2((0,T)\times\Omega)^d$ and the estimates hold*

$$\|m\|_{L^\infty(0,T;H_0^1(\Omega))} \leq C\left(\|f\|_{L^2((0,T)\times\Omega)} + \left\|m^0\right\|_{H_0^1(\Omega)}\right), \tag{2.12}$$

$$\|\Delta m\|_{L^2((0,T)\times\Omega)} \leq C\left(\|f\|_{L^2((0,T)\times\Omega)} + \left\|m^0\right\|_{H_0^1(\Omega)}\right). \tag{2.13}$$

The same statement is true for the problem (2.10)–(2.11), (2.8) in the case $\gamma = 1/2$.

Proof. In both cases (i.e. $\gamma \geq 1/2$) we construct the solution of the differential inclusion

$$\frac{\partial m}{\partial t} + \partial\mathcal{I}_\gamma(m) \ni f \tag{2.14}$$

with the functional $\mathcal{I}_\gamma : L^2(\Omega) \to [0,+\infty]$ given by

$$\mathcal{I}_\gamma(m) := \frac{D^2}{2}\int_\Omega |\nabla m|^2 \,\mathrm{d}x + \frac{1}{2\gamma}\int_\Omega |m|^{2\gamma}\,\mathrm{d}x, \qquad \text{if } m \in H_0^1(\Omega)\cap L^{2\gamma}(\Omega),$$

and $\mathcal{I}_\gamma(m) := +\infty$ otherwise. It can be easily checked that for $\gamma \geq 1/2$ the functional $\mathcal{I}_\gamma$ is proper with dense domain, strictly convex and lower semicontinuous on $H_0^1(\Omega)$. By the Rockafellar theorem [13], the Fréchet subdifferential $\partial\mathcal{I}_\gamma(m)$ is a maximal monotone operator and the standard theory [2] then provides the existence of a unique solution $m \in L^2(0,T;H_0^1(\Omega))^d \cap L^{2\gamma}((0,T)\times\Omega)^d$ of (2.14). Clearly, for $\gamma > 1/2$, m is the unique weak solution of (2.7)–(2.8). For $\gamma = 1/2$, m is the so-called *slow solution*, meaning that the velocity $\frac{\mathrm{d}m}{\mathrm{d}t}$ is the element of minimal norm in $\partial\mathcal{R}(m)$, i.e.,

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\mathrm{argmin}\{\|u\|_{H_0^1(\Omega)}\,; u \in \partial\mathcal{R}(m)\}.$$

Therefore, m is a weak solution of (2.10)–(2.11).

To prove the higher regularity estimates (2.12), (2.13), we use (formally, but easily justifiable) Δm as a test function, which after integration by parts leads to

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |\nabla m|^2\,\mathrm{d}x + D^2\int_\Omega |\Delta m|^2\,\mathrm{d}x - \int_\Omega (|m|^{2(\gamma-1)}m)\cdot\Delta m\,\mathrm{d}x = \int_\Omega f\cdot\Delta m\,\mathrm{d}x. \tag{2.15}$$

Then, denoting $\varphi(m) := |m|^{2(\gamma-1)}m$, we have

$$-\int_\Omega (|m|^{2(\gamma-1)}m)\cdot\Delta m\,\mathrm{d}x = -\int_\Omega \varphi(m)\cdot\Delta m\,\mathrm{d}x = \int_\Omega \nabla m\cdot D\varphi(m)\nabla m\,\mathrm{d}x \tag{2.16}$$

with

$$D\varphi(m) = |m|^{2(\gamma-1)}\left(2(\gamma-1)\frac{m}{|m|}\otimes\frac{m}{|m|} + I\right).$$

Clearly, $D\varphi(m)$ is a nonnegative matrix for $\gamma \geq 1/2$, so that the term (2.16) is nonnegative. The identity (2.15) together with a standard density argument gives directly the required regularity and the estimates (2.12), (2.13).

■

The rest of the proof of existence of solutions of the regularized problem (2.5)–(2.6) is identical to Section 2 of [7]. For the limit $\varepsilon \to 0$, we only need to provide the following result for the case $\gamma = 1/2$.

Lemma 3. *Let $m^k \to m$ strongly in $L^1((0,T)\times\Omega)$ as $k\to\infty$, and denote $h^k := r(m^k)$ with r given by (2.11). Then there exists $h \in \partial\mathcal{R}(m)$ such that, for a whileuence,*

$$h^k \rightharpoonup^* h \qquad \text{weakly* in } L^\infty((0,T)\times\Omega) \text{ as } k\to\infty.$$

Proof: Because $h^k = r(m^k)$ is uniformly bounded in $L^\infty((0,T)\times\Omega)$, there exists a subsequence, still denoted by h^k, converging to $h \in L^\infty((0,T)\times\Omega)$ weakly*. Due to the strong convergence of m^k in L^1, there exists a subsequence converging almost everywhere to m. Consequently, h^k converges to $m/|m|$ almost everywhere on $\{m \neq 0\}$. On $\{m = 0\}$, we have $|h| \leq 1$, so that $h \in \partial\mathcal{R}(m)$ defined by (2.2).

■

Note that in the case $\gamma = 1/2$, due to Lemma 3, we only obtain weak solutions of the system (1.6), (2.1). We conjecture that m is in fact a slow solution of (2.1), i.e., that it solves

$$\partial_t m - D^2\Delta m - c^2(m\cdot\nabla p[m])\nabla p[m] = r(m)$$

with $r(m)$ given by (2.11).

Remark 1. *The proof of local in time existence of mild solutions (Theorem 3 of [7]) carries over to the case $1/2 < \gamma < 1$ without modifications. However, the proof of uniqueness of mild solutions by a contraction mapping argument requires $\gamma \geq 1$ and it is not clear how to adapt it for values of γ less than one.*

3 Analysis in the 1d setting

Much more can be proved about the system (1.6)–(1.7) in the spatially one dimensional setting. Then, and without loss of generality, we can consider it on the interval $\Omega := (0,1)$. The system reads

$$-\partial_x\big(\partial_x p + m^2\partial_x p\big) = S, \tag{3.1}$$

$$\partial_t m - D^2\partial_{xx}^2 m - c^2(\partial_x p)^2 m + |m|^{2(\gamma-1)}m = 0, \tag{3.2}$$

Additionally, throughout this section we assume $S > 0$ a.e. on $(0,1)$, and for mathematical convenience we prescribe the mixed boundary conditions for p,

$$\partial_x p(0) = 0, \qquad p(1) = 0,$$

and homogeneous Neumann boundary condition for m. Then, integrating (3.1) with respect to x, we obtain

$$(1+m^2)\partial_x p = -\int_0^x S(y)\,\mathrm{d}y.$$

Denoting $B(x) := \int_0^x S(y)\,\mathrm{d}y$, we have

$$\partial_x p = -\frac{B(x)}{1+m^2}, \tag{3.3}$$

so that the system (3.1)–(3.2) is rewritten as

$$\partial_t m - D^2\partial_{xx}^2 m = \left(\frac{c^2 B(x)^2}{(1+m^2)^2} - |m|^{2(\gamma-1)}\right)m. \tag{3.4}$$

3.1 Extinction of solutions for $-1 \le \gamma \le 1$ and small sources

We show that if the source term S is small enough in a suitable sense, then solutions of (3.4) converge to zero, either in infinite time for $1/2 \le \gamma \le 1$, or in finite time for $-1 \le \gamma < 1/2$. In the latter case it means that the solutions can only exist on finite time intervals, since the algebraic term $|m|^{2(\gamma-1)}m$ is singular at $m = 0$ and solutions of (3.4) cannot be extended beyond the point where they reach zero.

Lemma 4. *Let $D \ge 0$, $-1 \le \gamma \le 1$, $m^0 \in L^\infty(0,1)$ and $c\,\|B\|_{L^\infty(0,1)} < Z_\gamma$ with*

$$Z_\gamma := \frac{2}{\gamma+1}\left(\frac{1-\gamma}{1+\gamma}\right)^{\frac{\gamma-1}{2}}, \qquad Z_1 := 1. \tag{3.5}$$

Without loss of generality, let $\inf_{x\in(0,1)} |m^0(x)| > 0$, and m be a weak solution of (3.4) with homogeneous Neumann boundary conditions and $m(t=0) = m^0$.

Then, for $1/2 \le \gamma \le 1$, $\|m\|_{L^1(0,1)}$ converges to zero as $t \to \infty$. For $-1 \le \gamma < 1/2$ there exists a finite break-down time $T_0 > 0$ such that $\inf_{x\in(0,1)} |m(T_0,x)| = 0$.

Proof. For $m > 0$ we define the positive function

$$h_\gamma(m) := m^{2(\gamma-1)}\,(1+m^2)^2. \tag{3.6}$$

It can be easily shown that for all $-1 \le \gamma \le 1$,

$$\inf_{0<m<\infty} h_\gamma(m) = Z_\gamma^2 > 0.$$

Consequently, the assumption $c\|B\|_{L^\infty(0,1)} < Z_\gamma$ implies

$$\frac{c^2\|B\|^2_{L^\infty(0,1)}}{(1+m^2)^2} - |m|^{2(\gamma-1)} < 0 \qquad \text{for all } m \in \mathbb{R}.$$

As a consequence of the maximum principle for (3.4), we have the a-priori bound

$$\|m(t)\|_{L^\infty(0,1)} \le M := \|m^0\|_{L^\infty(0,1)} \qquad \text{for all } t \ge 0.$$

Now, we can conclude that there exists a $\delta > 0$ such that

$$\frac{c^2\|B\|^2_{L^\infty(0,1)}}{(1+m^2)^2} - |m|^{2(\gamma-1)} < -\delta \qquad \text{for } |m| \le M. \tag{3.7}$$

As long as $\inf_{x\in(0,1)} |m(t)| > 0$, we can multiply (3.4) with $\operatorname{sign}(m)$ and integrate over $\Omega = (0,1)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 |m|\,\mathrm{d}x = D^2\int_0^1 (\partial^2_{xx}m)\operatorname{sign}(m)\,\mathrm{d}x + \int_0^1 \left(\frac{c^2B^2|m|}{(1+m^2)^2} - |m|^{2\gamma-1}\right)\mathrm{d}x.$$

On the one hand, the Kato inequality [3] for the first term of the right-hand side yields

$$D^2\int_0^1 (\partial^2_{xx}m)\operatorname{sign}(m)\,\mathrm{d}x \le D^2\int_0^1 \partial^2_{xx}|m|\,\mathrm{d}x = D^2\Big[\partial_x|m|\Big]_{x=0}^1 = D^2\Big[(\partial_x m)\operatorname{sign}(m)\Big]_{x=0}^1 = 0,$$

where the boundary term vanishes due to the homogeneous Neumann boundary conditions. Now, for $1/2 \leq \gamma \leq 1$, (3.7) implies

$$\int_0^1 \left(\frac{c^2 B^2 |m|}{(1+m^2)^2} - |m|^{2\gamma-1} \right) \mathrm{d}x < -\delta \int_0^1 |m| \, \mathrm{d}x,$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_0^1 |m| \, \mathrm{d}x < -\delta \int_0^1 |m| \, \mathrm{d}x$$

and by Gronwall lemma we conclude exponential convergence of $\|m\|_{L^1(0,1)}$ (or, due to the maximum principle, any L^q-norm of m with $q < \infty$) to zero as $t \to \infty$.

For $-1 \leq \gamma < 1/2$, from (3.7) and the behaviour near $m \approx 0$ it follows that there exists a $\widetilde{\delta} > 0$ such that

$$\frac{c^2 \|B^2\|_{L^\infty(0,1)} |m|}{(1+m^2)^2} - |m|^{2\gamma-1} < -\widetilde{\delta} \qquad \text{for } |m| \leq M.$$

Therefore, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_0^1 |m| \, \mathrm{d}x \leq -\widetilde{\delta}$$

and conclude the result with $T_0 < \left\|m^0\right\|_{L^1(0,1)} / \widetilde{\delta}$.

■

Remark 2. *For $\gamma < -1$, there exists a unique positive solution m_b of the equation*

$$\frac{c^2 B^2}{(1+m^2)^2} - |m|^{2(\gamma-1)} = 0$$

for each $cB > 0$. Consequently, the claim of Lemma 4 cannot be extended to the case $\gamma < -1$ in a straightforward way. It can be done under the smallness assumption on the initial datum $|m^0(x)| < m_b(x)$ for all $x \in (0,1)$, but we will skip the technical details here.

3.2 Nonlinear stability analysis for $D = 0$

In Section 6.1 of [7] we studied the nonlinear asymptotic stability of the 1d network formation system with $D = 0$ and $\gamma \geq 1$. We now extend that analysis to values $\gamma \geq 1/2$.

Setting $D = 0$ in (3.4), we obtain

$$\partial_t m = \left(\frac{c^2 B(x)^2}{(1+m^2)^2} - |m|^{2(\gamma-1)} \right) m, \tag{3.8}$$

which we interpret as a family of ODEs for $m = m(t)$ with the parameter x. Assuming that $S > 0$ on $(0,1)$, we have $B(x) > 0$ on $(0,1)$.

Clearly, $m = 0$ is a steady state for (3.8); with $\gamma = 1/2$ we interpret $m/|m| = 0$ for $m = 0$. To find nonzero steady states, we solve the algebraic equation

$$\frac{c^2 B(x)^2}{(1+m^2)^2} - |m|^{2(\gamma-1)} = 0,$$

in other words, we look for the roots of $h_\gamma(m) - c^2 B^2(x) = 0$ with h_γ given by (3.6). We distinguish the cases:

- $\gamma > 1$: The ODE (3.8) has three stationary points: *unstable* $m_0 = 0$ and *stable* $\pm m_s$. Therefore, the asymptotic steady state for (3.8) subject to the initial datum $m^0 = m^0(x)$ on $(0,1)$ is $m_s(x)\text{sign}(m^0(x))$.
- $\gamma = 1$:
  - If $c|B(x)| > 1$, then there are three stationary points, *unstable* $m_0 = 0$ and *stable* $\pm\sqrt{c|B(x)| - 1}$.
  - If $c|B(x)| \le 1$, then there is the only *stable* stationary point $m = 0$.

  Thus, the solution of (3.8) subject to the initial datum $m^0 = m^0(x)$ on $(0,1)$ converges to the asymptotic steady state $\chi_{\{c|B(x)|>1\}}(x)\text{sign}(m^0(x))\sqrt{c|B(x)| - 1}$.
- For $1/2 \le \gamma < 1$ the picture depends on the size of $c|B(x)|$ relative to Z_γ defined in (3.5).
  - If $c|B(x)| > Z_\gamma$, then (3.8) has five stationary points, *stable* $m_0 = 0$, *unstable* $\pm m_u$ and *stable* $\pm m_s$, with $0 < m_u < m_s$.
  - If $c|B(x)| = Z_\gamma$, then zero is a *stable* stationary point and there are two symmetric nonzero stationary points (attracting from $\pm\infty$ and repulsing towards zero).
  - If $c|B(x)| < Z_\gamma$, then there is the only *stable* stationary point $m = 0$.

Remark 3. *The above asymptotic stability result for the case $1/2 \le \gamma < 1$ shows that, at least in the case $D = 0$, the assumption $c\,\|B\|_{L^\infty(0,1)} < Z_\gamma$ of Lemma 4 is optimal.*

4 Stationary solutions in the multidimensional setting for $D = 0$

In the multidimensional setting we are able to construct *pointwise* stationary solutions of (1.6)–(1.7). Regarding the number of possible solutions, we obtain the same picture as in the previous Section 3.2. However, we are not able to provide a stability analysis.

We denote $u := (I + m \otimes m)\nabla p$, so that (1.6) gives

$$-\nabla \cdot u = S$$

and

$$\nabla p = (I + m \otimes m)^{-1} u = \left(I - \frac{m \otimes m}{1 + |m|^2}\right) u. \tag{4.1}$$

The activation term $c^2(m \cdot \nabla p)\nabla p$ in (1.7) is then expressed in terms of u as

$$c^2(m \cdot \nabla p)\nabla p = c^2 \frac{m \cdot u}{1 + |m|^2}\left(I - \frac{m \otimes m}{1 + |m|^2}\right) u.$$

Therefore, stationary solutions of (1.6)–(1.7) with $D = 0$ satisfy

$$c^2 \frac{m \cdot u}{1 + |m|^2} u = \left(c^2 \frac{(m \cdot u)^2}{(1 + |m|^2)^2} + |m|^{2(\gamma-1)} \right) m. \tag{4.2}$$

Clearly, $m(x) = 0$ is a solution for any $u \in \mathbb{R}^d$. On the other hand, if $m(x) \neq 0$, then there exists a nonzero scalar $\beta(x) \in \mathbb{R} \setminus \{0\}$ such that $m(x) = \beta(x)u(x)$. Denoting $z := \beta(x)|u(x)|$ and inserting into (4.2) gives

$$\frac{c^2|u|^2}{1 + z^2} = \frac{c^2|u|^2 z^2}{(1 + z^2)^2} + |z|^{2(\gamma-1)},$$

which further reduces to

$$c|u| = |z|^{\gamma-1}(1 + z^2). \tag{4.3}$$

We now distinguish the cases:

- For $\gamma > 1$ the equation (4.3) has exactly one positive solution $z > 0$ for every $|u| > 0$.
- For $\gamma = 1$ the equation (4.3) has exactly one positive solution $z > 0$ for every $|u| > 1/c$ and no positive solutions for $|u| \leq 1/c$.
- For $1 > \gamma \geq 1/2$ (in fact for $\gamma > -1$, but we discard the values of $\gamma < 1/2$), if $c|u| > Z_\gamma$ with Z_γ given by (3.5), there exist exactly two positive solutions $z_1, z_2 > 0$ of (4.3) for every $c|u| > 0$. If $c|u| = Z_\gamma$, there is one positive solution $z > 0$, and if $c|u| < Z_\gamma$, (4.3) has no solutions.

Let us recall that in [7] we considered stationary solutions (m_0, p_0) of (1.6)–(1.7) in the case $D = 0$, $\gamma > 1$. These are constructed by fixing measurable disjoint sets $\mathcal{A}_+ \subseteq \Omega$, $\mathcal{A}_- \subseteq \Omega$ and setting

$$m_0(x) := \left(\chi_{\mathcal{A}_+}(x) - \chi_{\mathcal{A}_-}(x) \right) c^{\frac{1}{\gamma-1}} |\nabla p_0(x)|^{\frac{2-\gamma}{\gamma-1}} \nabla p_0(x), \tag{4.4}$$

where $p_0 \in H_0^1(\Omega)$ solves the nonlinear Poisson equation

$$-\nabla \cdot \left[\left(1 + c^{\frac{2}{\gamma-1}} |\nabla p_0(x)|^{\frac{2}{\gamma-1}} \chi_{\mathcal{A}_+ \cup \mathcal{A}_-}(x) \right) \nabla p_0(x) \right] = S, \tag{4.5}$$

subject to - say - homogeneous Dirichlet boundary condition. The steady states $p_0 \in H_0^1(\Omega) \cap W_0^{1,2\gamma/(\gamma-1)}(\mathcal{A}_+ \cup \mathcal{A}_-)$ were found as the unique minimizers of the uniformly convex and coercive functional

$$\mathcal{F}_\gamma[p] := \frac{1}{2} \int_\Omega |\nabla p|^2 \, \mathrm{d}x + c^{\frac{2}{\gamma-1}} \frac{\gamma - 1}{2\gamma} \int_{\mathcal{A}_+ \cup \mathcal{A}_-} |\nabla p|^{\frac{2\gamma}{\gamma-1}} \, \mathrm{d}x - \int_\Omega pS \, \mathrm{d}x,$$

see Theorem 6 in [7]. Let us remark that the linearized stability analysis performed in Section 6.2 of [7] implies that in the case $D = 0$, $\gamma > 1$ the *linearly* stable (in the sense of Gâteaux derivative) networks fill up the whole domain due to the necessary condition $\mathrm{meas}(\mathcal{A}_+ \cup \mathcal{A}_-) = \mathrm{meas}(\Omega)$. In the 1d case, the nonlinear stability analysis of Section 3.2 above implies that the same holds also for the (nonlinearly) stable stationary solution. On the other hand, for $\gamma = 1$ the stationary solution m_0 *must* vanish on the set $\{x \in \Omega;\ |u(x)| \leq 1/c\}$. We shall return to this case below.

4.1 Stationary solutions in the multidimensional setting for $D = 0$, $1/2 \le \gamma < 1$

Inserting the formula $m(x) = \frac{z(x)}{|u(x)|}u(x)$ into (4.1) gives

$$\begin{aligned} \nabla p(x) &= u(x) \qquad \text{for } m(x) = 0, \\ &= \frac{u(x)}{1 + z(x)^2} \qquad \text{for } m(x) \neq 0. \end{aligned}$$

Consequently, we choose mutually disjoint measurable sets $\mathcal{A}_0$, $\mathcal{A}_1$, $\mathcal{A}_2$ such that $\mathcal{A}_0 \cup \mathcal{A}_1 \cup \mathcal{A}_2 = \Omega$, and construct the stationary pressure gradient as

$$\nabla p(x) = a(x, |u(x)|^2)u(x)$$

with

$$a(x, r) = \chi_{\{r<\overline{Z}\}} + \chi_{\{r\ge\overline{Z}\}} \left(\chi_{\mathcal{A}_0}(x) + \frac{\chi_{\mathcal{A}_1}(x)}{1 + z_1(r)^2} + \frac{\chi_{\mathcal{A}_2}(x)}{1 + z_2(r)^2} \right), \tag{4.6}$$

where we denoted $\overline{Z} := Z_\gamma^2/c^2$, and $z_1(r)$, $z_2(r)$ are the two positive solutions of (4.3) with $r = |u|^2$, i.e.,

$$c^2 r = |z(r)|^{2(\gamma-1)}(1 + z(r)^2)^2. \tag{4.7}$$

We denote by $z_1(r)$ the branch of solutions of (4.7) that is decreasing in r, while $z_2(r)$ is increasing. Consequently,

$$\text{range}(z_1) = (0, z(\overline{Z})], \qquad \text{range}(z_2) = [z(\overline{Z}), \infty),$$

where we denoted $z(\overline{Z}) := z_1(\overline{Z}) = z_2(\overline{Z})$.

We assume $\int_\Omega S(x)\,\mathrm{d}x = 0$ and prescribe the homogeneous Neumann boundary condition for p,

$$\nabla p \cdot \nu = 0 \qquad \text{on } \partial\Omega,$$

where ν is the outer normal vector to $\partial\Omega$. We perform the Helmholtz decomposition of u as

$$u = \nabla\varphi + \operatorname{curl} U,$$

where φ solves

$$\begin{aligned} -\Delta\varphi &= S \qquad \text{in } \Omega, \\ \nabla\varphi \cdot \nu &= 0 \qquad \text{on } \partial\Omega. \end{aligned}$$

The identity $\operatorname{curl} \nabla p = \operatorname{curl}(a(x, |u|)u) = 0$ gives the equation

$$\operatorname{curl}\left[a(x, |\nabla\varphi + \operatorname{curl} U|^2)(\nabla\varphi + \operatorname{curl} U)\right] = 0, \tag{4.8}$$

subject to the boundary condition $\operatorname{curl} U \cdot \nu = 0$ on $\partial\Omega$.

We define $A(x, r) := \int_0^r a(x, s)\,\mathrm{d}s \ge 0$ and for given $\varphi = \varphi(x)$ the functional

$$\mathcal{I}(U) := \frac{1}{2}\int_\Omega A(x, |\nabla\varphi + \operatorname{curl} U|^2)\,\mathrm{d}x. \tag{4.9}$$

It is easily checked that (4.8) is the Euler-Lagrange equation corresponding to critical points of $\mathcal{I}$.

We will now check whether $\mathcal{I}$ is convex. For any fixed vector $\xi \in \mathbb{R}^d$ have

$$\begin{aligned}\frac{1}{2}\xi \cdot D^2_{ww} A(x, |\nabla\varphi + w|^2)\xi &= \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \partial^2_{w_i w_j} A(x, |\nabla\varphi + w|^2)\xi_i \xi_j \\ &= 2\frac{\partial^2 A}{\partial r^2}(x, |\nabla\varphi + w|^2)(\xi \cdot (\nabla\varphi + w))^2 + \frac{\partial A}{\partial r}(x, |\nabla\varphi + w|^2)|\xi|^2,\end{aligned}$$

where $\frac{\partial A}{\partial r} = \frac{\partial A}{\partial r}(x, r)$ is the derivative of A with respect to the second variable. Therefore, convexity of $\mathcal{I}(U)$ is equivalent to the condition

$$2\frac{\partial^2 A}{\partial r^2}(x, |v|^2)(\xi \cdot v)^2 + \frac{\partial A}{\partial r}(x, |v|^2)|\xi|^2 \geq 0 \qquad \text{for all } v,\, \xi \in \mathbb{R}^d. \tag{4.10}$$

Using the decomposition $\xi = \lambda v + v^\perp$ with $\lambda = \frac{\xi \cdot v}{|v|^2}$ and $v^\perp \cdot v = 0$ gives

$$\lambda^2 |v|^2 \left(2\frac{\partial^2 A}{\partial r^2}(x, |v|^2)|v|^2 + \frac{\partial A}{\partial r}(x, |v|^2)\right) + \frac{\partial A}{\partial r}(x, |v|^2)|v^\perp|^2 \geq 0.$$

Consequently, (4.10) is equivalent to the conditions

$$\frac{\partial A}{\partial r}(x, r) \geq 0, \qquad 2\frac{\partial^2 A}{\partial r^2}(x, r)r + \frac{\partial A}{\partial r}(x, r) \geq 0$$

for all $x \in \Omega$, $r > 0$. Note that $\frac{\partial A}{\partial r}(x, r) = a(x, r) \geq 0$ due to (4.6), so that we only need to verify the second condition. We choose a compactly supported nonnegative test function $\psi \in C_0^\infty(\Omega, [0, \infty))$ and integrate by parts to obtain

$$\int_0^\infty \int_\Omega \left(2\frac{\partial^2 A}{\partial r^2}(x, r)r + \frac{\partial A}{\partial r}(x, r)\right)\psi(x, r)\,\mathrm{d}x\,\mathrm{d}r = -\int_0^\infty \int_\Omega a(x, r)\left(2r\frac{\partial \psi}{\partial r} + \psi\right)\,\mathrm{d}x\,\mathrm{d}r.$$

Inserting for $a = a(x, r)$ the expression (4.6), we calculate

$$\begin{aligned}-\int_0^\infty \int_\Omega a(x, r)\left(2r\frac{\partial \psi}{\partial r} + \psi\right)\,\mathrm{d}x\,\mathrm{d}r &= \int_0^{\overline{Z}} \int_\Omega \psi(x, r)\,\mathrm{d}x\,\mathrm{d}r + \int_{\overline{Z}}^\infty \int_{\mathcal{A}_0} \psi(x, r)\,\mathrm{d}x\,\mathrm{d}r - 2\overline{Z}\int_{\mathcal{A}_1 \cup \mathcal{A}_2} \psi(x, \overline{Z})\,\mathrm{d}x \\ &+ \int_{\overline{Z}}^\infty \int_{\mathcal{A}_1} \frac{\psi(x, r)}{1 + z_1(r)^2}\,\mathrm{d}x\,\mathrm{d}r + \int_{\overline{Z}}^\infty \int_{\mathcal{A}_2} \frac{\psi(x, r)}{1 + z_2(r)^2}\,\mathrm{d}x\,\mathrm{d}r \\ &+ \frac{2\overline{Z}}{1 + z_1(\overline{Z})^2}\int_{\mathcal{A}_1} \psi(x, \overline{Z})\,\mathrm{d}x + \frac{2\overline{Z}}{1 + z_2(\overline{Z})^2}\int_{\mathcal{A}_2} \psi(x, \overline{Z})\,\mathrm{d}x \\ &+ \frac{4}{c^2}\int_0^{z_1(\overline{Z})} \int_{\mathcal{A}_1} s^{2\gamma - 1}\psi(x, z_1^{-1}(s))\,\mathrm{d}s - \frac{4}{c^2}\int_{z_2(\overline{Z})}^\infty \int_{\mathcal{A}_2} s^{2\gamma - 1}\psi(x, z_2^{-1}(s))\,\mathrm{d}s.\end{aligned}$$

Note that even if we set $\mathcal{A}_2 := \emptyset$, the third term of the right-hand side cannot be, in general, balanced by the other ones. Consequently, we do not have convexity of $\mathcal{I}(U)$.

To check coercivity, we take $r > \overline{Z}$, then

$$A(x, r) = \overline{Z} + (r - \overline{Z})\chi_{\mathcal{A}_0}(x) + \chi_{\mathcal{A}_1}(x)\int_{\overline{Z}}^r \frac{\mathrm{d}s}{1 + z_1(s)^2} + \chi_{\mathcal{A}_2}(x)\int_{\overline{Z}}^r \frac{\mathrm{d}s}{1 + z_2(s)^2}.$$

With (4.7) we have for both branches z_1, z_2,

$$c^2 r = |z|^{1(\gamma+1)} \left(1 + 2z^{-2} + z^{-4}\right).$$

Consequently, the increasing branch $z_2(r) \to \infty$ when $r \to \infty$ and $c^2 r \sim z_2(r)^{2(1+\gamma)}$, so that

$$\frac{1}{1 + z_2(r)^2} \sim c^{-\frac{2}{1+\gamma}} r^{-\frac{1}{1+\gamma}}$$

and

$$\int^r \frac{\mathrm{d}s}{1 + z_2(s)^2} \sim r^{\frac{\gamma}{1+\gamma}}.$$

Noting that $r = |u|^2 = |\mathrm{curl}\ U + \nabla\varphi|^2$, the energy estimate gives (at least) control of $|\mathrm{curl}\ U|^{\frac{2\gamma}{1+\gamma}}$. For $1/2 \le \gamma < 1$ this gives the range $2/3 \le \frac{2\gamma}{1+\gamma} < 1$ which is not enough to obtain usable coercivity estimates. Thus, the existence of stationary points of the functional $\mathcal{I}$ remains open for $1/2 \le \gamma < 1$, however, the corresponding variational formulation (4.9) can be used as an alternative method for numerical simulations.

4.2 Stationary solutions in the multidimensional setting for $D = 0$, $\gamma = 1$

In the case $\gamma = 1$, the stationary version of (1.7) with $D = 0$ reads

$$c^2(\nabla p_0 \otimes \nabla p_0) m_0 = m_0,$$

i.e., m_0 is either the zero vector or an eigenvector of the matrix $c^2(\nabla p_0 \otimes \nabla p_0)$ with eigenvalue 1. The spectrum of $c^2(\nabla p_0 \otimes \nabla p_0)$ consists of zero and $c^2|\nabla p_0|^2$, so that $m_0 \neq 0$ is only possible if $c^2|\nabla p_0|^2 = 1$. Therefore, for every stationary solution there exists a measurable function $\lambda = \lambda(x)$ such that

$$m_0(x) = \lambda(x) \chi_{\{c^2|\nabla p_0|^2 = 1\}}(x) \nabla p_0(x)$$

and p_0 solves the highly nonlinear Poisson equation

$$-\nabla \cdot \left[\left(1 + \frac{\lambda(x)^2}{c^2} \chi_{\{c^2|\nabla p_0|^2 = 1\}}(x) \right) \nabla p_0 \right] = S$$

subject to the homogeneous Dirichlet boundary condition $p_0 = 0$ on $\partial\Omega$.

A simple consideration suggests that *stable* stationary solutions of (1.6)–(1.7) with $D = 0$ should be constructed as

$$\begin{aligned} -\nabla \cdot \left[(1 + a(x)^2) \nabla p_0 \right] &= S, \qquad p_0 \in H_0^1(\Omega), && (4.11) \\ c^2 |\nabla p_0(x)|^2 &\le 1, \qquad \text{a.e. on } \Omega, && (4.12) \\ a(x)^2 \left[c^2 |\nabla p_0(x)|^2 - 1 \right] &= 0, \qquad \text{a.e. on } \Omega, && (4.13) \end{aligned}$$

for some measurable function $a^2 = a(x)^2$ on Ω which is the Lagrange multiplier for the condition (4.12). This condition follows from the nonpositivity of the eigenvalues of the matrix $c^2(\nabla p_0 \otimes \nabla p_0) - I$, which is heuristically a necessary condition for linearized stability of the stationary solution of (1.6)–(1.7) with $D = 0$. The function $\lambda = \lambda(x)$ can be chosen as $\lambda(x) := c a(x)$.

4.2.1 Variational formulation

We claim that solutions of (4.11)–(4.13) are minimizers of the energy functional

$$\mathcal{J}[p] := \int_\Omega \left(\frac{|\nabla p|^2}{2} - Sp \right) \mathrm{d}x \tag{4.14}$$

on the set $\mathcal{M} := \{p \in H_0^1(\Omega), c^2|\nabla p|^2 \leq 1 \text{ a.e. on } \Omega\}$.

Lemma 5. *Let $S \in L^2(\Omega)$. There exists a unique minimizer of the functional* (4.14) *on the set $\mathcal{M}$. It is the unique weak solution of the problem* (4.11)–(4.13) *with homogeneous Dirichlet boundary conditions on Ω and with $a \in L^2(\Omega)$.*

Proof. The functional $\mathcal{J}$ is convex and, due to the Poincaré inequality, coercive on $H_0^1(\Omega)$. Therefore, a unique minimizer $p_0 \in H_0^1(\Omega)$ exists on the closed, convex set $\mathcal{M}$. Clearly, (4.11)–(4.13) is the Euler-Lagrange system corresponding to this constrained minimization problem, so that p_0 is its weak solution. Moreover, using p_0 as a test function and an application of the Poincaré inequality yields

$$\begin{aligned} \int_\Omega (1+a^2)|\nabla p_0|^2 \,\mathrm{d}x &= \int_\Omega S p_0 \,\mathrm{d}x \\ &\leq \int_\Omega |\nabla p_0|^2 \,\mathrm{d}x + C \int_\Omega S^2 \,\mathrm{d}x. \end{aligned}$$

With (4.13) we have then

$$\int_\Omega a^2 \,\mathrm{d}x = c^2 \int_\Omega a^2 \nabla |p_0|^2 \,\mathrm{d}x \leq c^2 C \int_\Omega S^2 \,\mathrm{d}x,$$

so that $a \in L^2(\Omega)$.

Next, we prove that any weak solution $p_0 \in H_0^1(\Omega)$, $a^2 \in L^1(\Omega)$ of (4.11)–(4.13) is a minimizer of (4.14) on the set $\mathcal{M}$. Indeed, we consider any $q \in \mathcal{M}$ and use $(p_0 - q)$ as a test function for (4.11),

$$\int_\Omega (1+a^2) \nabla p_0 \cdot \nabla (p_0 - q) \,\mathrm{d}x = \int_\Omega S(p_0 - q) \,\mathrm{d}x.$$

The Cauchy-Schwarz inequality for the term $\nabla p_0 \cdot \nabla q$ gives

$$\frac{1}{2} \int_\Omega (1+a^2)|\nabla p_0|^2 \,\mathrm{d}x \leq \frac{1}{2} \int_\Omega (1+a^2)|\nabla q|^2 \,\mathrm{d}x + \int_\Omega (p_0 - q) S \,\mathrm{d}x.$$

Moreover, (4.13) gives $a^2|\nabla p_0|^2 = a^2/c^2$, and with $|\nabla q| \leq 1/c^2$ we have

$$\int_\Omega \left(\frac{|\nabla p_0|^2}{2} - p_0 S \right) \mathrm{d}x + \frac{1}{2c^2} \int a^2 \,\mathrm{d}x \leq \int_\Omega \frac{|\nabla q|^2}{2} - qS \,\mathrm{d}x + \frac{1}{2c^2} \int a^2 \,\mathrm{d}x,$$

so that $\mathcal{J}[p_0] \leq \mathcal{J}[q]$.

Finally, let $p_i \in H_0^1(\Omega)$, $a_i \in L^2(\Omega)$, $i = 1, 2$, be two weak solutions of (4.11)–(4.13). We take the difference of (4.11) for p_1 and p_2 and test by $(p_1 - p_2)$:

$$\int_\Omega \left[(1+a_1^2) \nabla p_1 - (1+a_2^2) \nabla p_2 \right] \cdot (\nabla p_1 - \nabla p_2) \,\mathrm{d}x = 0.$$

We use the Cauchy-Schwarz inequality for

$$\int_\Omega (a_1^2+a_2^2)(\nabla p_1\cdot\nabla p_2)\,\mathrm{d}x \le \frac{1}{2}\int_\Omega (a_1^2+a_2^2)|\nabla p_1|^2\,\mathrm{d}x + \frac{1}{2}\int_\Omega (a_1^2+a_2^2)|\nabla p_2|^2\,\mathrm{d}x \le \frac{1}{c^2}\int_\Omega (a_1^2+a_2^2),$$

where the second inequality comes from (4.12). Consequently, we have

$$\int_\Omega |\nabla p_1-\nabla p_2|^2\,\mathrm{d}x + \int_\Omega a_1^2|\nabla p_1|^2 + a_2^2|\nabla p_2|^2 \le \frac{1}{c^2}\int_\Omega (a_1^2+a_2^2).$$

Finally, using (4.13) we obtain

$$\int_\Omega |\nabla p_1-\nabla p_2|^2\,\mathrm{d}x \le 0$$

and conclude that $p_1 = p_2$ a.e. on Ω.

■

Remark 4. *The gradient constrained variational problem* (4.14) *was studied in [5] as a model for twisting of an elastic-plastic cylindrical bar. There it was shown that the unique solution has $C^{1,1}$-regularity in Ω; see also [16, 17].*

4.2.2 A penalty method for $D = 0$, $\gamma = 1$

Solutions of (4.11)–(4.13) can also be constructed via a penalty approximation. Although the variational formulation used in the previous section is a short, effective and elegant way to prove existence of solutions, we provide the alternative penalty method here, since it provides approximations of the solution and since we find the related analytical techniques interesting on their own. For the following we assume Ω to be the unit cube $(0,1)^d$ and prescribe periodic boundary conditions on $\partial\Omega$ to discard of cumbersome boundary terms.

We consider the penalized problem

$$-\nabla\cdot\left[\left(1+\frac{(|\nabla p_\varepsilon|^2-1/c^2)_+}{\varepsilon}\right)\nabla p_\varepsilon\right] = S, \qquad p_\varepsilon\in\overline{H}^1(\Omega), \tag{4.15}$$

where $A_+ := \max(A,0)$ denotes the positive part of A and $\overline{H}^1(\Omega) = \left\{u\in H^1_{\mathrm{per}}(\Omega),\ \int_\Omega u\,\mathrm{d}x = 0\right\}$. Here $H^1_{\mathrm{per}}(\Omega)$ denotes the space of $H^1_{\mathrm{loc}}(\mathbb{R}^d)$-functions with $(0,1)^d$-periodicity.

Theorem 2. *For any $S\in L^2(\Omega)$ with $\int_\Omega S(x)\,\mathrm{d}x = 0$ there exists a unique weak solution $p\in\overline{H}^1(\Omega)$ of* (4.15).

Proof: The functional $\mathcal{F}_\varepsilon : \overline{H}^1(\Omega)\to\mathbb{R}\cup\{\infty\}$,

$$\mathcal{F}_\varepsilon[p] := \frac{1}{2}\int_\Omega |\nabla p|^2\,\mathrm{d}x + \frac{1}{4\varepsilon}\int_\Omega (|\nabla p|^2-1/c^2)_+^2\,\mathrm{d}x - \int_\Omega pS\,\mathrm{d}x,$$

is uniformly convex and coercive on $\overline{H}^1(\Omega)$. The classical theory (see, e.g., [6]) provides the existence of a unique minimizer $p_\varepsilon\in\overline{H}^1(\Omega)$ of $\mathcal{F}_\varepsilon$, which is a weak solution of the corresponding Euler-Lagrange

equation (4.15). The uniqueness of solutions follows from the monotonicity of the function $F : \mathbb{R}^d \to \mathbb{R}^d$, $F(x) = (|x|^2 - 1/c^2)_+ x$.

■

We will prove convergence of a subsequence of $\{p_\varepsilon\}_{\varepsilon>0}$, solutions of (4.15), towards a solution of (4.11)–(4.13) as $\varepsilon \to 0$. We introduce the notation

$$a_\varepsilon := \frac{(|\nabla p_\varepsilon|^2 - 1/c^2)_+}{\varepsilon}. \tag{4.16}$$

Theorem 3. *Let $S \in H^1(\Omega)$ with $\int_\Omega S(x)\,\mathrm{d}x = 0$, $d \le 3$ and $(p_\varepsilon)_{\varepsilon>0} \subset \overline{H}^1(\Omega)$ be a family of solutions of (4.15) constructed in Theorem 2, and $(a_\varepsilon)_{\varepsilon>0} \subset L^2(\Omega)$ given by (4.16). Then there exist a subsequence of $(p_\varepsilon, a_\varepsilon)$ and $p \in \overline{H}^1(\Omega)$, $a^2 \in L^2(\Omega)$ such that, as $\varepsilon \to 0$,*

- *$\nabla p_\varepsilon \to \nabla p$ strongly in $L^2(\Omega)$ and strongly in $L^4(\Omega)$.*
- *$(1 + a_\varepsilon)\nabla p_\varepsilon \rightharpoonup (1 + a^2)\nabla p$ weakly in $L^1(\Omega)$, so that (4.11) is satisfied in the weak sense.*
- *$(|\nabla p_\varepsilon|^2 - 1/c^2)_+ = \varepsilon a_\varepsilon \to 0$ strongly in $L^2(\Omega)$, so that $(|\nabla p|^2 - 1/c^2)_+ = 0$ and (4.12) is satisfied a.e.*
- *$a_\varepsilon \left[|\nabla p_\varepsilon|^2 - 1/c^2\right] = \varepsilon a_\varepsilon^2 \to 0$ strongly in $L^1(\Omega)$, and $a_\varepsilon \left[|\nabla p_\varepsilon|^2 - 1/c^2\right] \rightharpoonup a \left[|\nabla p|^2 - 1/c^2\right]$ weakly in $L^1(\Omega)$ if $d \le 3$, so that (4.13) is satisfied a.e.*

The proof of the above Theorem is based on the following a priori estimates:

Lemma 6. *The family $(p_\varepsilon)_{\varepsilon>0}$ constructed in Theorem 2 is uniformly bounded in $L^\infty(\Omega)$.*

Proof: This is a direct consequence of the maximum principle for (4.15).

■

Lemma 7. *Let $S \in H^1(\Omega)$. Then the solutions p_ε of (4.15) satisfy*

$$\int_\Omega |\nabla^2 p_\varepsilon|^2 dx \le C \left\|\nabla S\right\|^2_{L^2(\Omega)}, \qquad \int_\Omega (1 + a_\varepsilon)|\nabla^2 p_\varepsilon|^2\,\mathrm{d}x \le C \left\|\nabla S\right\|^2_{L^2(\Omega)}.$$

Proof. We use the short-hand notation $\partial_i := \partial_{x_i}$ and denote $p_i := \partial_i p_\varepsilon$ and $p_{ij} := \partial^2_{ij} p_\varepsilon$. Moreover, we set $w_\varepsilon(x) := |\nabla p_\varepsilon(x)|$. We use the Bernstein method and differentiate (4.15) with respect to x_j:

$$-\partial_i \left[(1 + a_\varepsilon)\, p_{ij}\right] - \frac{1}{\varepsilon}\partial_i \left[\chi_{\{(cw-1)_+\}}(\partial_j w^2) p_i\right] = S_j.$$

Then we multiply by p_j and integrate by parts

$$\int_\Omega (1 + a_\varepsilon)\, p_{ij}^2\,\mathrm{d}x + \frac{1}{\varepsilon}\int_{\{(cw-1)_+\}} (\partial_j w^2) p_i p_{ij}\,\mathrm{d}x = \int_\Omega S_j p_j\,\mathrm{d}x.$$

Now we use the identity $2p_i p_{ij} = (\partial_j w^2)$, so that the second term of the left-hand side becomes

$$\frac{1}{2\varepsilon}\int_{\{(cw-1)_+\}} (\partial_j w^2)^2\,\mathrm{d}x = \frac{1}{2\varepsilon}\int_\Omega \left[\partial_j (w^2 - 1/c^2)_+\right]^2\,\mathrm{d}x.$$

Therefore, we have

$$\int_\Omega (1+a_\varepsilon)\, p_{ij}^2 \,\mathrm{d}x + \frac{1}{2\varepsilon}\int_\Omega \left[\partial_j (w^2 - 1/c^2)_+\right]^2 \,\mathrm{d}x = \int_\Omega S_j p_j \,\mathrm{d}x.$$

Using $a_\varepsilon \geq 0$ and the nonnegativity of the second term of the left-hand side, we write

$$\frac{1}{2}\int_\Omega (1+a_\varepsilon)\, p_{ij}^2 \,\mathrm{d}x + \frac{1}{2}\int_\Omega p_{ij}^2 \,\mathrm{d}x \leq \int_\Omega (1+a_\varepsilon)\, p_{ij}^2 \,\mathrm{d}x \leq \int_\Omega S_j p_j \,\mathrm{d}x.$$

The claim follows by using a Cauchy-Schwarz and Poincaré inequality (with constant C_Ω) in the right-hand side,

$$\int_\Omega S_j p_j \,\mathrm{d}x \leq \frac{1}{2\delta}\int_\Omega S_j^2 \,\mathrm{d}x + \frac{\delta}{2}\int_\Omega p_j^2 \,\mathrm{d}x \leq \frac{1}{2\delta}\int_\Omega |\nabla S|^2 \,\mathrm{d}x + \frac{\delta C_\Omega}{2}\int_\Omega |\nabla^2 p|^2 \,\mathrm{d}x,$$

and choosing δ such that $\frac{\delta C_\Omega}{2} < 1/2$,

$$\frac{1}{2}\int_\Omega (1+a_\varepsilon)\, p_{ij}^2 \,\mathrm{d}x + C\int_\Omega p_{ij}^2 \,\mathrm{d}x \leq \frac{1}{2\delta}\int_\Omega |\nabla S|^2 \,\mathrm{d}x,$$

with $C = \frac{1}{2} - \frac{\delta C_\Omega}{2} > 0$.

■

Lemma 8. *The solutions p_ε of* (4.15) *and a_ε given by* (4.16) *satisfy*

$$\int_\Omega (1+a_\varepsilon)^2 |\nabla^2 p_\varepsilon|^2 \,\mathrm{d}x \leq \|S\|^2_{L^2(\Omega)}, \qquad \int_\Omega |\nabla a_\varepsilon \cdot \nabla p_\varepsilon|^2 \,\mathrm{d}x \leq 2\,\|S\|^2_{L^2(\Omega)}.$$

Proof: We again use the short-hand notation from the previous proof, and, moreover, denote $a_i := \partial_i a_\varepsilon$. We take the square of (4.15),

$$\begin{aligned}
\int_\Omega S^2 \,\mathrm{d}x &= \int_\Omega \left[\sum_{i=1}^d \partial_i[(1+a_\varepsilon)p_i]\right]\left[\sum_{j=1}^d \partial_j[(1+a_\varepsilon)p_j]\right] \mathrm{d}x \\
&= \sum_{i=1}^d\sum_{j=1}^d \int_\Omega [\partial_j[(1+a_\varepsilon)p_i]][\partial_i(1+a_\varepsilon)p_j]] \,\mathrm{d}x \\
&= \sum_{i=1}^d\sum_{j=1}^d \int_\Omega [a_j p_i + (1+a_\varepsilon)p_{ij}][a_i p_j + (1+a_\varepsilon)p_{ij}]] \,\mathrm{d}x \\
&= \left(\sum_{i=1}^d \int_\Omega a_i p_i \,\mathrm{d}x\right)^2 + \sum_{i=1}^d\sum_{j=1}^d \int_\Omega (1+a_\varepsilon)p_{ij}(a_j p_i + a_i p_j) \,\mathrm{d}x + \int_\Omega (1+a_\varepsilon)^2 |\nabla^2 p_\varepsilon|^2 \,\mathrm{d}x.
\end{aligned}$$

In the middle term of the last line we use the identity $\partial_{x_i}|\nabla p|^2 = 2\sum_{j=1}^d p_j p_{ij}$, so that

$$\sum_{i=1}^d\sum_{j=1}^d \int_\Omega (1+a_\varepsilon)p_{ij}(a_j p_i + a_i p_j) \,\mathrm{d}x = \sum_{i=1}^d \int_\Omega (1+a_\varepsilon)a_i \partial_{x_i}|\nabla p|^2 \,\mathrm{d}x.$$

Morever, denoting $w := |\nabla p|^2$, we realize that $a_\varepsilon(w) = \frac{(w-1/c^2)_+}{\varepsilon}$ is a nondecreasing function of w, so that $a_i \partial_{x_i} |\nabla p|^2 = (\partial_{x_i} a_\varepsilon(w))(\partial_{x_i} w) = a'_\varepsilon(w)(\partial_{x_i} w)^2 \geq 0$. Consequently, the middle term is nonnegative and we have

$$\int_\Omega (1+a_\varepsilon)^2 |\nabla^2 p_\varepsilon|^2 \,\mathrm{d}x \leq \int_\Omega S^2 \,\mathrm{d}x.$$

Now, knowing that $(1+a_\varepsilon)\Delta p_\varepsilon$ is bounded in $L^2(\Omega)$ due to Lemma 7, and expanding the derivatives in (4.15),

$$(1+a_\varepsilon)\Delta p_\varepsilon + \nabla p_\varepsilon \cdot \nabla a_\varepsilon = -S,$$

we conclude

$$\|\nabla p_\varepsilon \cdot \nabla a_\varepsilon\|_{L^2(\Omega)} \leq \|S\|_{L^2(\Omega)} + \|(1+a_\varepsilon)\Delta p_\varepsilon\|_{L^2(\Omega)} \leq 2\,\|S\|_{L^2(\Omega)}\,.$$

■

Lemma 9. *The sequence $(a_\varepsilon)_{\varepsilon>0}$ defined in (4.16) is uniformly bounded in $L^2(\Omega)$.*

Proof: We multiply (4.15) by $a_\varepsilon p_\varepsilon$ and integrate by parts. This gives

$$\int_\Omega (1+a_\varepsilon) a_\varepsilon |\nabla p_\varepsilon|^2 \,\mathrm{d}x + \int_\Omega (1+a_\varepsilon) p_\varepsilon \nabla p_\varepsilon \cdot \nabla a_\varepsilon \,\mathrm{d}x = \int_\Omega S p_\varepsilon a_\varepsilon \,\mathrm{d}x.$$

We write the first term as

$$\begin{aligned}\int_\Omega (1+a_\varepsilon) a_\varepsilon |\nabla p_\varepsilon|^2 \,\mathrm{d}x &= \int_\Omega (1+a_\varepsilon) a_\varepsilon (|\nabla p_\varepsilon|^2 - 1/c^2)\,\mathrm{d}x + \frac{1}{c^2}\int_\Omega (1+a_\varepsilon) a_\varepsilon \,\mathrm{d}x \\ &= \varepsilon \int_\Omega (1+a_\varepsilon) a_\varepsilon^2 \,\mathrm{d}x + \frac{1}{c^2}\int_\Omega (1+a_\varepsilon) a_\varepsilon \,\mathrm{d}x.\end{aligned}$$

Due to the nonnegativity of the first term, we have

$$\begin{aligned}\frac{1}{c^2}\int_\Omega a_\varepsilon^2 \,\mathrm{d}x &\leq \frac{1}{c^2}\int_\Omega (1+a_\varepsilon) a_\varepsilon \,\mathrm{d}x \leq \int_\Omega (1+a_\varepsilon) a_\varepsilon |\nabla p_\varepsilon|^2 \,\mathrm{d}x \\ &\leq \|S\|_{L^2(\Omega)} \|p_\varepsilon\|_{L^\infty(\Omega)} \|a_\varepsilon\|_{L^2(\Omega)} + \|\nabla p_\varepsilon \cdot \nabla a_\varepsilon\|_{L^2(\Omega)} \|p_\varepsilon\|_{L^\infty(\Omega)} \|1+a_\varepsilon\|_{L^2(\Omega)} \\ &\leq \frac{1}{2\delta}\|S\|^2_{L^2(\Omega)} \|p_\varepsilon\|^2_{L^\infty(\Omega)} + \frac{\delta}{2}\|a_\varepsilon\|^2_{L^2(\Omega)} \\ &\quad + \frac{1}{2\delta}\|\nabla p_\varepsilon \cdot \nabla a_\varepsilon\|^2_{L^2(\Omega)} \|p_\varepsilon\|^2_{L^\infty(\Omega)} + \delta\left(|\Omega|^2 + \|a_\varepsilon\|^2_{L^2(\Omega)}\right)\end{aligned}$$

for any $\delta > 0$. Then, an application of Lemmata 6 and 8 gives a constant $C > 0$ such that

$$\frac{1}{c^2}\|a_\varepsilon\|^2_{L^2(\Omega)} \leq C(1+\delta^{-1}) + \frac{3\delta}{2}\|a_\varepsilon\|^2_{L^2(\Omega)}$$

and choosing $\delta > 0$ small enough, we conclude.

■

Now we are ready to prove Theorem 3:

Proof: From Lemma 7 we conclude that $\nabla^2 p_\varepsilon$ is uniformly bounded in $L^2(\Omega)$, so, in $d \leq 3$ spatial dimensions, ∇p_ε converges strongly in $L^4(\Omega)$ to ∇p due to the compact Sobolev embedding.

Since a_ε is bounded in $L^2(\Omega)$ by Lemma 9, there exists a weakly converging subsequence to $a^2 \in L^2(\Omega)$. Thus, due to the strong convergence of ∇p_ε, the product $(1 + a_\varepsilon)\nabla p_\varepsilon$ converges weakly in $L^1(\Omega)$ to $(1 + a^2)\nabla p$.

Clearly, $\varepsilon a_\varepsilon \to 0$ strongly in $L^2(\Omega)$. Moreover, due to the inequality $|a_+ - b_+| \leq |a - b|$, we have

$$0 \leq \int_\Omega \left| (|\nabla p_\varepsilon|^2 - 1/c^2)_+ - (|\nabla p|^2 - 1/c^2)_+ \right| \,\mathrm{d}x \leq \int_\Omega \left| |\nabla p_\varepsilon|^2 - |\nabla p|^2 \right| \,\mathrm{d}x,$$

so that the strong convergence of ∇p_ε in $L^2(\Omega)$ implies

$$\varepsilon a_\varepsilon = (|\nabla p_\varepsilon|^2 - 1/c^2)_+ \to (|\nabla p|^2 - 1/c^2)_+ \qquad \text{strongly in } L^1(\Omega).$$

Consequently, $(|\nabla p|^2 - 1/c^2)_+ = 0$ a.e.

Clearly, $\varepsilon a_\varepsilon^2 \to 0$ strongly in $L^1(\Omega)$. The strong convergence of ∇p_ε in $L^4(\Omega)$ and weak convergence (of a subsequence of) a_ε in $L^2(\Omega)$ implies

$$\varepsilon a_\varepsilon^2 = a_\varepsilon \left[|\nabla p_\varepsilon|^2 - 1/c^2 \right] \rightharpoonup a^2 \left[|\nabla p|^2 - 1/c^2 \right] \qquad \text{weakly in } L^1(\Omega).$$

Consequently, $a^2 \left[|\nabla p|^2 - 1/c^2 \right] = 0$ a.e.

■

Finally, let us note that uniqueness of solutions of the system (4.11)–(4.13) was already proved in Lemma 5.

4.3 Stationary solutions via the variational formulation for $D = 0$, $1/2 \leq \gamma < 1$

Let us recall that in [7] we constructed stationary solutions (m_0, p_0) of (1.6)–(1.7) in the case $\gamma > 1$, $D = 0$ by using (4.4) and employing the variational formulation of the nonlinear Poisson equation (4.5). Clearly, this approach fails for $\gamma < 1$ due to the singularity of the term $|\nabla p_0(x)|^{\frac{2-\gamma}{\gamma-1}} \nabla p_0(x)$ at $|\nabla p_0| = 0$ and the resulting non-boundedness from below of the associated functional. However, stationary solutions can be constructed by "cutting off" small values of $|\nabla p_0|$. For simplicity, we set the activation parameter $c^2 := 1$ in this section.

We fix a measurable set $\mathcal{A} \subset \Omega$ and a constant $\alpha > 0$ to be specified later, and define the stationary solution of (1.6), (1.7) for $D = 0$:

$$m_0 = \chi_{\mathcal{A}} \chi_{\{|\nabla p_0| > \alpha\}} |\nabla p_0|^{\frac{2-\gamma}{\gamma-1}} \nabla p_0, \tag{4.17}$$

where p_0 solves

$$-\nabla \cdot [(1 + m_0 \otimes m_0)\nabla p_0] = S.$$

This is the Euler-Lagrange equation corresponding to the functional $\mathcal{F}_\alpha : H_0^1(\Omega) \to \mathbb{R}$,

$$\mathcal{F}_\alpha[p] = \int_\Omega \frac{|\nabla p|^2}{2} + \frac{\gamma - 1}{2\gamma} \chi_{\mathcal{A}}(x) \left(|\nabla p|^{\frac{2\gamma}{\gamma-1}} - \alpha^{\frac{2\gamma}{\gamma-1}} \right)_- \,\mathrm{d}x. \tag{4.18}$$

We examine the convexity of $\mathcal{F}$ in dependence on the value of α. Defining $F : \mathbb{R}^d \to \mathbb{R}$,

$$F(\xi) := \frac{|\xi|^2}{2} + \frac{\gamma - 1}{2\gamma} \left(|\xi|^{\frac{2\gamma}{\gamma-1}} - \alpha^{\frac{2\gamma}{\gamma-1}} \right)_- ,$$

we calculate the Hessian matrix

$$D^2 F(\xi) = F''(|\xi|) \frac{\xi \otimes \xi}{|\xi|^2} + F'(|\xi|) \left(\frac{I}{|\xi|} - \frac{\xi \otimes \xi}{|\xi|^3} \right).$$

This has the eigenvectors ξ and $\xi^\perp$ and a quick inspection reveals that F is convex as a function of ξ if and only if $F'(|\xi|) \geq 0$ and $F''(|\xi|) \geq 0$. Writing $r := |\xi|$, we have

$$\begin{aligned} F'(r) &= r + r^{\frac{\gamma+1}{\gamma-1}}, \qquad r > \alpha, \\ &= r, \qquad r < \alpha, \end{aligned}$$

and

$$\begin{aligned} F''(r) &= 1 + \delta(r - \alpha)\alpha^{\frac{\gamma+1}{\gamma-1}} + \frac{\gamma+1}{\gamma-1} r^{\frac{2}{\gamma-1}}, \qquad r \geq \alpha, \\ &= 1, \qquad r < \alpha. \end{aligned}$$

Thus, F is a uniformly convex function on $\mathbb{R}^d$ if and only if $\alpha > \alpha_\gamma$ with

$$\alpha_\gamma := \left(\frac{1 - \gamma}{1 + \gamma} \right)^{\frac{\gamma-1}{2}}. \tag{4.19}$$

Lemma 10. *Let $1/2 \leq \gamma < 1$ and $\alpha > \alpha_\gamma$ with α_γ given by* (4.19)*. Then the functional* (4.18) *is coercive and uniformly convex on $H_0^1(\Omega)$ and the unique minimizer p_0 with m_0 given by* (4.17) *is a stationary solution of* (1.6)–(1.7)*.*

Unfortunately, in the spatially one-dimensional case we are able to show that the above construction delivers the nonlinearly *unstable* steady states (as long as $m_0 \not\equiv 0$), so these solutions will never be observed in the long time limit of the system (1.6)–(1.7). For simplicity, we set $\Omega := (0, 1)$ and $\mathcal{A} = \emptyset$. Let us recall that the 1d problem with $D = 0$ reduces to the ODE family

$$\partial_t m = \left(\frac{B(x)^2}{(1 + m^2)^2} - |m|^{2(\gamma-1)} \right) m. \tag{4.20}$$

As calculated in Section 3.2, the nonlinearly *stable* stationary solution of (4.20) for $1/2 < \gamma < 1$ is

$$\begin{aligned} m(x) &= 0, \qquad \text{if } |B(x)| \leq Z_\gamma, \\ m(x) &\in \{0, \pm m_s(x)\}, \qquad \text{if } |B(x)| > Z_\gamma, \end{aligned}$$

where Z_γ is given by (3.5) and $m_s > 0$ is the *largest* solution of

$$\frac{B(x)^2}{(1 + m^2)^2} = |m|^{2(\gamma-1)},$$

with $B(x) = \int_0^x S(y)\,\mathrm{d}y > 0$. Note that for $|B(x)| > Z_\gamma$ the above algebraic equation has four solutions $\pm m_u$, $\pm m_s$ with $0 < m_u < m_s$, and m_u is the unstable, m_s stable solution for (4.20). Moreover, note that

$$Z_\gamma = \frac{2}{1+\gamma}\alpha_\gamma,$$

so that $Z_\gamma > \alpha_\gamma$.

Now, let (m_0, p_0) be the solution constructed in Lemma 10, i.e., p_0 is the unique minimizer of (4.18) and m_0 is given by (4.17). Clearly, to have a *nonzero stable* state $m_0 = m_0(x)$ for some $x \in \Omega$, it is necessary that

$$|B(x)| > Z_\gamma \qquad \text{and} \qquad |\partial_x p_0(x)| > \alpha > \alpha_\gamma.$$

Moreover, if $m_0(x) \neq 0$, we have the formulas

$$|m_0| = |\partial_x p_0|^{\frac{1}{\gamma-1}} \qquad \text{and} \qquad |B(x)| = |m_0|^{\gamma-1}(1+|m_0|^2) = |\partial_x p_0|\left(1 + |\partial_x p_0|^{\frac{2}{\gamma-1}}\right).$$

Thus, defining $f_\gamma(u) := u\left(1+u^{\frac{2}{\gamma-1}}\right)$ for $u > 0$, we have $|B(x)| = f_\gamma(|\partial_x p_0|)$. The function $f_\gamma(u)$ has a unique strict minimum on $\mathbb{R}^+$ at $u = \alpha_\gamma > 0$ and $f_\gamma(\alpha_\gamma) = Z_\gamma$. Since $\lim_{u\to 0} f_\gamma(u) = \lim_{u\to+\infty} f_\gamma(u) = +\infty$, for each $|B(x)| > Z_\gamma$ there exist $0 < u_1 < \alpha_\gamma < u_2$ such that $f_\gamma(u_1) = f_\gamma(u_2) = |B(x)|$. Clearly, u_1, u_2 correspond to the nonzero steady states m_u, m_s of (4.20), and since $|m_0| = |\partial_x p_0|^{\frac{1}{\gamma-1}}$ is a decreasing function of $u = |\partial_x p_0|$ and $m_u < m_s$, the *unstable* steady state m_u corresponds to u_2, while the *stable* steady state m_s corresponds to u_1. Since, by construction, we have to choose $u = |\partial_x p_0| > \alpha_\gamma$ in order to have $m_0(x) \neq 0$, we are in fact choosing u_2 and thus the *unstable* steady state m_u.

5 Numerical Method and Examples

The model has the ability to generate fascinating patterns and we illustrate this with numerical experiments performed in two space dimensions using a Galerkin framework. These interesting patterns show up if the diffusivity in the system is low, i.e. $D \ll 1$, and the pressure gradient is large. In this context let us also mention [1, 8], where numerical simulations for Eqs. (1.6)–(1.7) have been presented.

Furthermore, we want to demonstrate that the results of the analysis in one dimension are also relevant for the two dimensional setting. For instance, for $\gamma < 1/2$ we are interested in extinction in finite time of the solution, cf. Section 3.1, and for $1/2 \leq \gamma < 1$ we demonstrate instability of solutions constructed in Section 4.3.

5.1 Mixed variational formulation

Since we are interested in the case $D \ll 1$, it turns out to be useful to reformulate Eq. (1.7) as a mixed problem. Consequently, setting $\sigma = \nabla m$, we consider

$$\begin{aligned}
-\nabla\cdot[(rI + m\otimes m)\nabla p] &= S \quad \text{in } \Omega\times(0,T),\\
p &= 0 \quad \text{on } \Gamma\times(0,T),\\
\nu\cdot(rI + m\otimes m)\nabla p &= 0 \quad \text{on } \partial\Omega\setminus\Gamma\times(0,T),\\
\partial_t m - D^2\nabla\cdot\sigma &= c^2(\nabla p\otimes\nabla p)m - |m|^{2(\gamma-1)}m \quad \text{in } \Omega\times(0,T),\\
\sigma - \nabla m &= 0 \quad \text{in } \Omega\times(0,T),\\
m &= 0 \quad \text{on } \partial\Omega\times(0,T),
\end{aligned}$$

with $m(t=0) = m^0$ in Ω. Here, $\Gamma \subset \partial\Omega$ denotes the Dirichlet part of the boundary, and we denote by $H^1_{0,\Gamma}(\Omega) = \{p \in H^1(\Omega) : p_{|\Gamma} = 0\}$ the space of Sobolev functions vanishing on Γ. Additionally, we need the space $H(\mathrm{div}) = \{\mu \in L^2(\Omega)^2 : \nabla\cdot\mu \in L^2(\Omega)\}$. As a starting point for our Galerkin framework we consider the following weak formulation: Find $(p, m, \sigma) \in L^\infty(0,T; H^1_{0,\Gamma}(\Omega)) \times L^2(0,T; L^2(\Omega)^2) \times L^2(0,T; H(\mathrm{div})^2)$ such that

$$\begin{aligned}
\int_\Omega (rI + m\otimes m)\nabla p \cdot \nabla q \,\mathrm{d}x &= \int_\Omega Sq \,\mathrm{d}x,\\
\int_\Omega \partial_t m \cdot v \,\mathrm{d}x - \int_\Omega D^2\nabla\cdot\sigma\cdot v \,\mathrm{d}x &= \int_\Omega f_{\gamma,c}(m,\nabla p)\cdot v \,\mathrm{d}x,\\
\int_\Omega \sigma\cdot\mu \,\mathrm{d}x + \int_\Omega m\cdot\nabla\cdot\mu \,\mathrm{d}x &= 0,
\end{aligned}$$

for all $(q, v, \mu) \in H^1_{0,\Gamma}(\Omega) \times L^2(\Omega)^2 \times H(\mathrm{div})^2$, and $m(t=0) = m^0$ in Ω. Here, we use the abbreviation

$$f_{\gamma,c}(m,\nabla p) = c^2(\nabla p \otimes \nabla p)m - |m|_\rho^{2(\gamma-1)}m,$$

where $|m|_\rho = \sqrt{m_1^2 + m_2^2 + \rho}$ is a regularized absolute value with regularization parameter $\rho \geq 0$. Any strong solution (m, p) to (1.6)–(1.7) satisfying the above boundary conditions yields a solution to the flux based weak formulation in case $\rho = 0$ and $\gamma > 1/2$. Homogeneous Neumann boundary conditions for m result in homogeneous Dirichlet boundary conditions for $\sigma\cdot\nu$ on $\partial\Omega$, and the function space for σ has to be adapted accordingly.

5.2 Space discretization

To obtain a space discretization, we let $\{\mathcal{T}_h\}$ be a family of regular quasi-uniform triangulations of Ω with $h = \max_{T\in\mathcal{T}_h} h_T$ and $h_T = \sqrt{|T|}$ for all $T \in \mathcal{T}_h$. For the approximation of the pressure p we choose standard Lagrangian finite elements, i.e. continuous, piecewise linear functions

$$P_h = \{v_h \in C^0(\overline{\Omega}) : v_{h|T} \in \mathcal{P}_1(T)\ \forall T \in \mathcal{T}_h, v_{h|\Gamma} = 0\} \subset H^1_{0,\Gamma}(\Omega).$$

For the approximation of the conductance vector m we choose piecewise constant functions

$$M_h = \{v_h \in L^2(\Omega) : v_{h|T} \in \mathcal{P}_0(T)\ \forall T \in \mathcal{T}_h\},$$

and for the approximation of σ we choose lowest order Raviart-Thomas elements

$$V_h = \{v_h \in H(\mathrm{div}) : v_{h|T} \in RT_0(T) = \mathcal{P}_0(T) + x\mathcal{P}_0(T)\ \forall T \in \mathcal{T}_h\}.$$

The resulting Galerkin approximation is then as follows. Find $(p_h, m_h, \sigma_h) \in L^\infty(0,T; P_h) \times L^2(0,T; M_h^2) \times L^2(0,T; V_h^2)$ such that for a.e. $t \in (0,T]$

$$\begin{aligned}
\int_\Omega (rI + m_h(t)\otimes m_h(t))\nabla p_h(t)\cdot\nabla q_h \,\mathrm{d}x &= \int_\Omega Sq_h \,\mathrm{d}x,\\
\int_\Omega \partial_t m_h(t)\cdot v_h \,\mathrm{d}x - \int_\Omega D^2\nabla\cdot\sigma_h(t)\cdot v_h \,\mathrm{d}x &= \int_\Omega f_{\gamma,c}(m_h(t),\nabla p_h(t))\cdot v_h \,\mathrm{d}x,\\
\int_\Omega \sigma_h(t)\cdot\mu_h \,\mathrm{d}x + \int_\Omega m_h(t)\cdot\nabla\cdot\mu_h \,\mathrm{d}x &= 0,
\end{aligned}$$

for all $(q_h, v_h, \mu_h) \in P_h \times M_h^2 \times V_h^2$, and $m_h(t=0) = m_h^0$ in Ω. Here, m_h^0 denotes the L^2-projection of m^0 onto M_h. Assuming sufficient regularity of solutions, the method applied to the stationary problem is of first order in the $L^2(\Omega)$-norm and also first order in the $L^2(\Omega)$-norm for ∇p and σ. The $L^2(\Omega)$-projection of m is approximated with second order in $L^2(\Omega)$ if $f_{\gamma,c}(m, \nabla p) \in H^1(\Omega)$. Our analytical results do not provide such regularity; however, even for regular solutions the error estimates are in practice not very helpful since the constants depend on norms of derivatives of solutions which are locally very large in the small diffusion - large activation regime. For details on the approximation spaces, mixed finite elements and corresponding error estimates see for instance [4].

5.3 Time discretization

For the discretization of the time variable let $0 = t^0 < t^1 < \ldots < t^K = T$ denote a partition of $[0, T]$. By $m_h^k \approx m_h(t^k)$, $p_h^k \approx p_h(t^k)$ and $\sigma_h^k \approx \sigma_h(t^k)$, $0 \le k \le K$, we denote the corresponding approximation in time, which is obtained by solving the following implicit-explicit (IMEX) first-order Euler scheme

$$\int_\Omega (rI + m_h^k \otimes m_h^k) \nabla p_h^k \cdot \nabla q_h \,\mathrm{d}x = \int_\Omega S q_h \,\mathrm{d}x, \tag{5.1}$$

$$\int_\Omega m_h^{k+1} \cdot v_h - \delta^{k+1} D^2 \nabla \cdot \sigma_h^{k+1} \cdot v_h \,\mathrm{d}x = \int_\Omega \left(m_h^k + \delta^{k+1} f_{\gamma,c}(m_h^k, \nabla p_h^k) \right) \cdot v_h \,\mathrm{d}x, \tag{5.2}$$

$$\int_\Omega \sigma_h^{k+1} \cdot \mu_h \,\mathrm{d}x + \int_\Omega m_h^{k+1} \cdot \nabla \cdot \mu_h \,\mathrm{d}x = 0, \tag{5.3}$$

for all $(q_h, v_h, \mu_h) \in P_h \times M_h^2 \times V_h^2$, and $\delta^{k+1} = t^{k+1} - t^k$. We note that for $D = 0$ Eq. (1.7) is an ODE, and (5.2) amounts to an explicit Euler scheme for approximating $m_h(t)$ on each triangle $T \in \mathcal{T}_h$. In addition, there is no coupling between the different triangles in this case, and consequently no numerical diffusion is introduced into the system.

The discrete counterpart of the energy defined in (1.5) is defined as follows

$$\mathcal{E}_h(m_h^k) = \frac{1}{2} \int_\Omega D^2 |\sigma_h^k|^2 + \frac{|m_h^k|_\rho^{2\gamma}}{\gamma} + c^2 |m_h^k \cdot \nabla p_h^k|^2 + c^2 |\nabla p_k^h|^2 \,\mathrm{d}x.$$

Our main guideline for obtaining a stable scheme is to ensure that $\mathcal{E}_h(m_h^{k+1}) \le \mathcal{E}_h(m_h^k)$ for all $k \ge 0$, which is inspired by but weaker than (2.4). We choose an adaptive time-stepping according to the following rule: $\delta^1 = 1/(2c^2 \|\|\nabla p_h^0\|\|_\infty^2)$, $t_1 = t_0 + \delta^1$, $\delta^2 = \delta^1$. Let δ^k be given. If $\delta^k \in (1/(20c^2 \|\|\nabla p_h^k\|\|_\infty^2), 9/(10c^2 \|\|\nabla p_h^k\|\|_\infty^2))$ then $\delta^{k+1} = \delta^k$, otherwise set $\delta^{k+1} = 1/(2c^2 \|\|\nabla p_h^k\|\|_\infty^2)$, and $t_{k+1} = t^k + \delta^{k+1}$. Moreover, we let δ^k be sufficiently small. This choice of time-step is motivated by the solution of the ODE system

$$\tilde{m}_t = \begin{pmatrix} 0 & 0 \\ 0 & c^2 |\nabla p|^2 \end{pmatrix} \tilde{m}, \qquad \tilde{m}(0) = \tilde{m}_0,$$

which is obtained from Eq. (1.7) with $D = 0$ and no relaxation term through diagonalization. Assuming $\nabla p(t)$ does not depend on t, the solution is given by

$$\tilde{m}_1(t) = \tilde{m}_{0,1}, \quad \tilde{m}_2(t) = \exp(c^2 |\nabla p|^2 t) \tilde{m}_{0,2}.$$

Since $c^2 \nabla p_h^k \otimes \nabla p_h^k$ is positive semi-definite the explicit Euler method is unstable for all choices of δ^k. Stability might however be retained through the relaxation term as soon as $c^2 \nabla p_h^k \otimes \nabla p_h^k - |m_h^k|^{2(\gamma-1)} I$

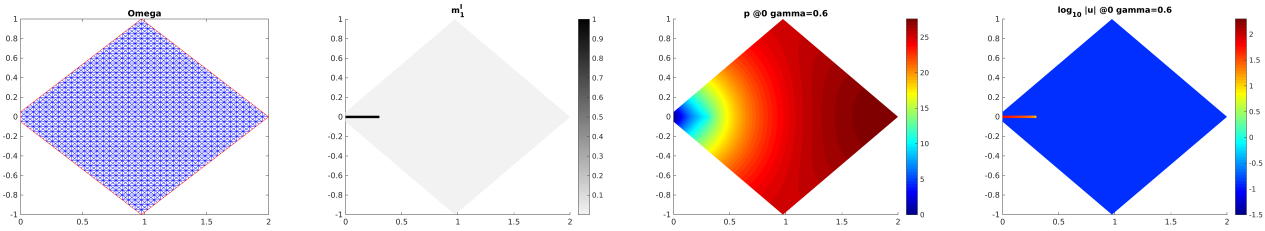

Figure 1: From left to right: Triangulation $\mathcal{T}_h$ of Ω with $1,678$ vertices and $3,196$ triangles; initial datum m_1^0, while $m_2^0 = 0$; initial pressure p_h^0; decadic logarithm of the absolute value of the initial velocity $|u_h^0|$. Note that p_h^0 and $|u_h^0|$ are the same for all values of γ.

is negative definite, i.e. if $c^2|\nabla p_h^k|^2 < |m_h^k|^{2(\gamma-1)}$; cf. Section 4.2. Besides this stability issue there is an additional linearization error by treating ∇p_h^k explicitly. Hence, if ∇p_h^k is changing rapidly, then δ^k should be sufficiently small to obtain a reasonable accuracy. A detailed investigation of stable and accurate time-stepping schemes is however out of the scope of this paper and is left for further research; let us mention [11, 15] for IMEX schemes in the context of reaction-diffusion equations.

5.4 Setup

As a computational domain we consider a diamond shaped two-dimensional domain with one edge cut, see Figure 1. We use a refined triangulation with $102,905$ vertices and $204,544$ triangles, which corresponds to $h \approx 0.0032$. The Dirichlet boundary is defined as $\Gamma = \{x \in \mathbb{R}^2 : x_1 = 0\} \cap \partial\Omega$. If not stated otherwise, we let

$$S = 1, \quad r = \frac{1}{10}, \quad c = 50, \quad D = \frac{1}{1000}, \quad \gamma = \frac{1}{2}, \quad \rho = 10^{-12},$$

where ρ is the regularization parameter defined in Section 5.1, and define the initial datum as

$$m_1^0(x) = \begin{cases} 1, & x \le 0.3 \text{ and } |y| \le 0.0125, \\ 0, & \text{else}, \end{cases} \qquad m_2^0 = 0.$$

The main quantity of interest is the discrete velocity defined as

$$u_h^k = (rI + m_h^k \otimes m_h^k)\nabla p_h^k,$$

see also Section 4. The initial velocity u_h^0 and the initial pressure p_h^0 do not depend on γ or D, and they are depicted in Figure 1. Since the numerical simulation is computationally expensive, we could not compute a stationary state in many examples below. However, in order to indicate that the presented solutions are near a stationary state, we define the stationarity measures

$$\mathcal{E}_{h,t}^k = \frac{\mathcal{E}_h(m_h^k) - \mathcal{E}_h(m_h^{k-1})}{\delta^k} \quad \text{and} \quad m_{h,t}^k = \frac{\|m_h^k - m_h^{k-1}\|_{L^2(\Omega)}}{\delta^k}.$$

Furthermore, we define the quantity

$$s_k = \frac{\|u_h^k\|_{L^2(\Omega)}}{\|u_h^k\|_{L^1(\Omega)}},$$

which measures the sparsity of the network. In order to demonstrate the dependence of the solution on the different parameters in the system we first present some simulations for varying parameter values.

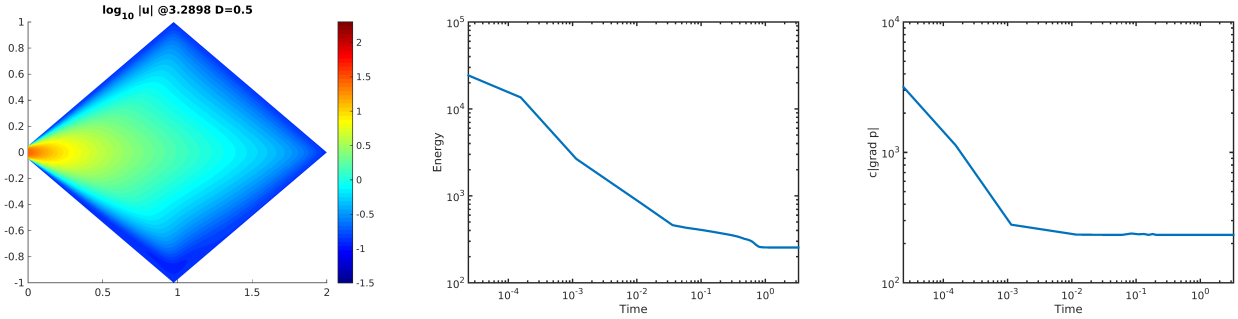

Figure 2: Stationary velocity $|u_h^k|$ for $\gamma = \frac{1}{2}$ and $D = \frac{1}{2}$ in a Log_{10}-scale, and corresponding evolution of $\mathcal{E}_h(m_h^k)$. The stationary state is reached for $t = 3.2898$.

5.5 Varying D

The proliferation of the network and its structure is crucially influenced by diffusion. In the limit of vanishing diffusion $D = 0$ the support of the conductance vector cannot grow. If diffusion is too large, interesting patterns will not show up in the stationary network. In the following we investigate the influence of different values for $D \in \{\frac{1}{2}, \frac{1}{10}, \frac{1}{100}, \frac{1}{1000}\}$ on the network formation, while keeping $\gamma = 1/2$ fixed. For $D \geq 1/10$, the obtained velocities are dominated by diffusion and no fine scale structures appear in the network, see Figures 2 and 3. For $D = 1/100$ the resulting velocity is still diffusive, but some large scale structure is visible, see Figure 4. Decreasing the diffusion coefficient even further to $D = 1/1000$, the network builds fine scale structures, see Figure 5. We note that the velocity is not near a stationary state here. For this very small $D = 1/1000$, we have to be careful in interpreting the results. Our simulations have shown a strong mesh dependence for this case, which is not apparent for $D \geq 1/2$. We are not able to fully explain this behavior, but the comparison on different meshes for large diffusion and moderate values of c, which makes in turn $c\nabla p$ moderate, suggest that the mesh is too coarse to be able to resolve the diffusion process properly in the presence of strong activation. Here, one should use finer meshes to resolve this issue, which however also leads to prohibitively long computation times. A modification of the existing scheme to cope with this issue is left to further research.

Nonetheless, we believe that the velocities presented here are qualitatively correct, as they structurally show the right behavior, i.e. the smaller the diffusion D the finer the scales in the network are, and the higher the sparsity index s_k is, see also Table 1. Let us remark that the closer γ is to $1/2$, the sparser the structures should be in a stationary state, which complies with the well-known fact that L^1-norm minimization promotes sparse solutions. Furthermore, even though the results in Figure 5 are quantitatively very different, they possess qualitatively the same properties; namely the thickness of the primary, secondary and tertiary branches. Let us again emphasize that the results of Figure 5 are far from being stationary, and the networks are likely to change their structure when further evolving.

5.6 Varying γ

In order to demonstrate the dependence of the network formation process on the relaxation term, we let $\gamma \in \{\frac{3}{5}, \frac{3}{4}, 1, \frac{3}{2}, 2\}$; for $\gamma = \frac{1}{2}$ see Section 5.5. For $\gamma \in \{\frac{3}{2}, 2\}$ the results are depicted in Figure 9 and Figure 10. From the evolution of the energies and from Table 2 we may conclude that for these two values of γ we are near a stationary state, and that for $\gamma > 1$ no fine scale structures built up. For

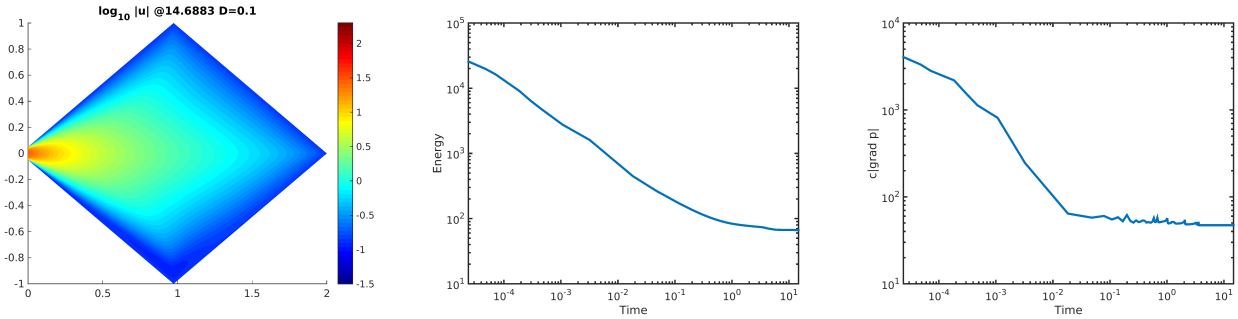

Figure 3: Near stationary velocity $|u_h^k|$ for $\gamma = \frac{1}{2}$ and $D = \frac{1}{10}$ in a Log_{10}-scale, and corresponding evolution of $\mathcal{E}_h(m_h^k)$ and $\|c|\nabla p_h^k|\|_{L^\infty(\Omega)}$.

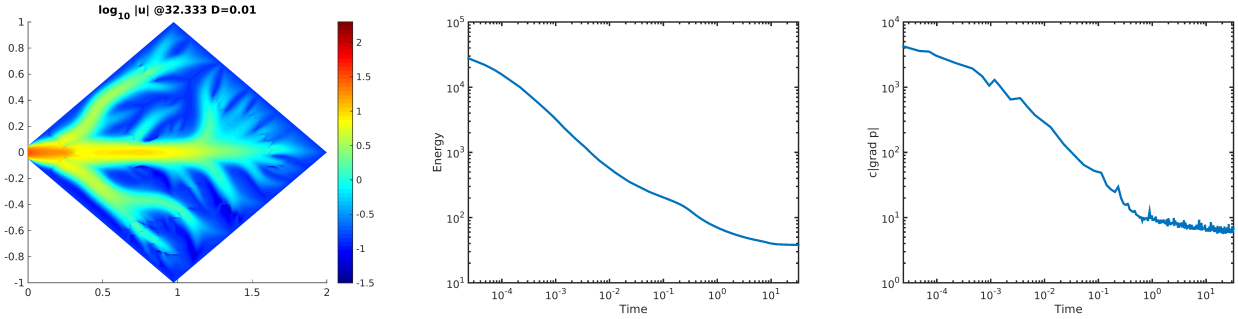

Figure 4: Near stationary velocity $|u_h^k|$ for $\gamma = \frac{1}{2}$ and $D = \frac{1}{100}$ in a Log_{10}-scale, and corresponding evolution of $\mathcal{E}_h(m_h^k)$ and $\|c|\nabla p_h^k|\|_{L^\infty(\Omega)}$.

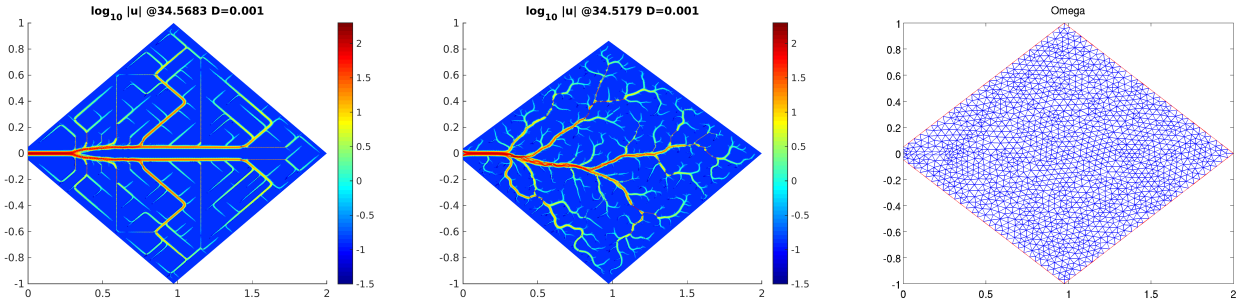

Figure 5: Left panel: Velocity $|u_h^k|$ for $\gamma = \frac{1}{2}$ and $D = \frac{1}{1000}$ in a Log_{10}-scale computed on a refined triangulation from the uniform grid shown in Figure 1 with $102,905$ vertices, cf. Section 5.4. Corresponding velocity (middle panel) computed on a refinement with $107,009$ vertices of the non-uniform grid shown in the right panel. For the refined non-uniform grid we have $\min h_T = 0.0017$ and $\max h_T = 0.0043$.

D	$\frac{1}{2}$	$\frac{1}{10}$	$\frac{1}{100}$
t^k	3.2898	14.6883	32.333
$\mathcal{E}_{h,t}^k$	-1.1×10^{-5}	-2.9×10^{-6}	-1.5×10^{-2}
$m_{h,t}^k$	6.2×10^{-7}	4.5×10^{-2}	2.8×10^{-1}
s_k	1.344153	1.348375	1.576061

Table 1: Stationarity and sparsity measures for different values of D and $\gamma = \frac{1}{2}$, see Section 5.5.

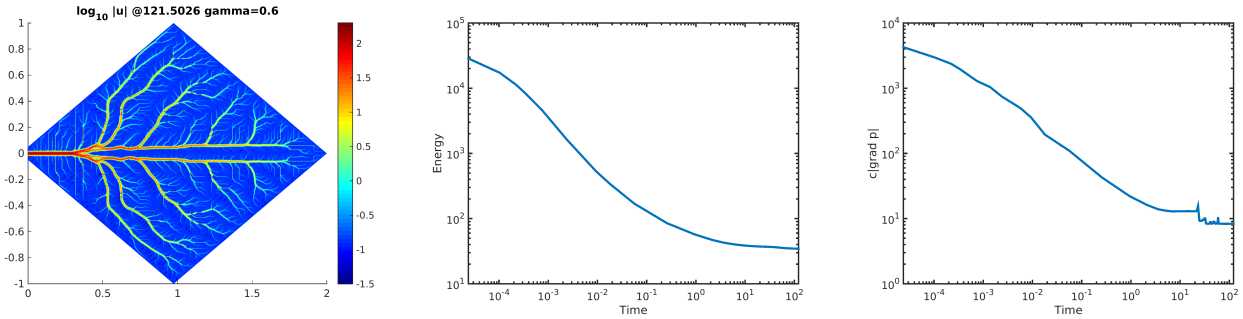

Figure 6: Near stationary velocity $|u_h^k|$ for $\gamma = \frac{3}{5}$, $D = \frac{1}{1000}$ in a Log_{10}-scale for different times, and corresponding evolution of the $\mathcal{E}_h(m_h^k)$ and $\|c|\nabla p|\|_{L^\infty(\Omega)}$ vs. time in a logarithmic scaling.

$\gamma = 1$, depicted in Figure 8, network structures appear for large times. This may also be indicated by the oscillating behavior of ∇p for larger times. The changes in energy are however already small. In view of Section 4.2, stable solutions should satisfy $c\||\nabla p_h^k|\|_{L^\infty(\Omega)} \leq 1$ in the limiting case $D = 0$. Here, $D = 1/1000$, and $c|\nabla p_h^k| \leq 2$ is in accordance with this analysis.

For $\gamma \in \{\frac{3}{5}, \frac{3}{4}\}$ we observe fine scale structures, which are depicted in Figure 6 and Figure 7. As remarked in the previous section, the network evolution is influenced by the underlying grid due to coarse discretization, very small diffusion, and very large activiation terms. Note that for small times we have $c\||\nabla p\|_{L^\infty(\Omega)} \approx 4000$, which enters quadratically in the activation term. The closer γ is to 1 the less the relaxation term promotes sparsity. This might explain that for $\gamma \geq \frac{3}{4}$ we see two branches originating from the Dirichlet boundary $\Gamma = \partial\Omega \cap \{x_1 = 0\}$. Since the pressure gradient is very large at the transition of Dirichlet to Neumann boundary, artificial conductance is created. Notice that these two branches do not appear for $\gamma \in \{\frac{1}{2}, \frac{3}{5}\}$. In this context let us mention that in several situations L^1-type minimization can be performed exactly by soft-shrinkage, where values below a certain threshold, which corresponds to δ^k here, are set to 0. Hence, small values of m, due to round-off or small diffusion, will less affect the evolution.

5.7 Unstable stationary solutions for $D = 0$ and $\frac{1}{2} \leq \gamma < 1$

In Section 4.3 we have constructed stationary solutions for $D = 0$ and $\frac{1}{2} \leq \gamma < 1$. In one dimension our stability analysis shows that these stationary states are not stable. In the following, we indicate that these stationary states are unstable also in two dimensions. To do so, we compute the minimizer of the functional $\mathcal{F}_\alpha$ defined in (4.18) with

$$\alpha = \alpha_\gamma = c^{-\frac{1}{4}} \Big(\frac{1-\gamma}{1+\gamma} \Big)^{\frac{\gamma-1}{2}},$$

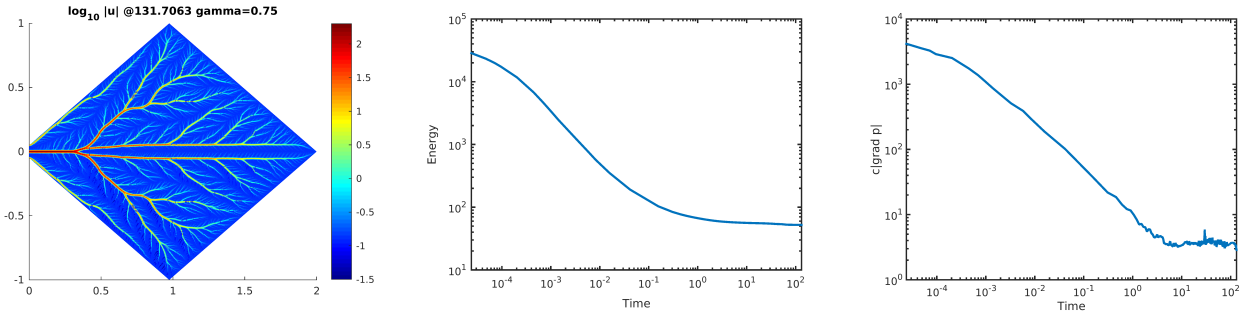

Figure 7: Velocity $|u_h^k|$ for $\gamma = \frac{3}{4}$, $D = \frac{1}{1000}$ in a Log_{10}-scale, and corresponding evolution of the $\mathcal{E}_h(m_h^k)$ and $\|c|\nabla p|\|_{L^\infty(\Omega)}$ vs. time in a logarithmic scaling.

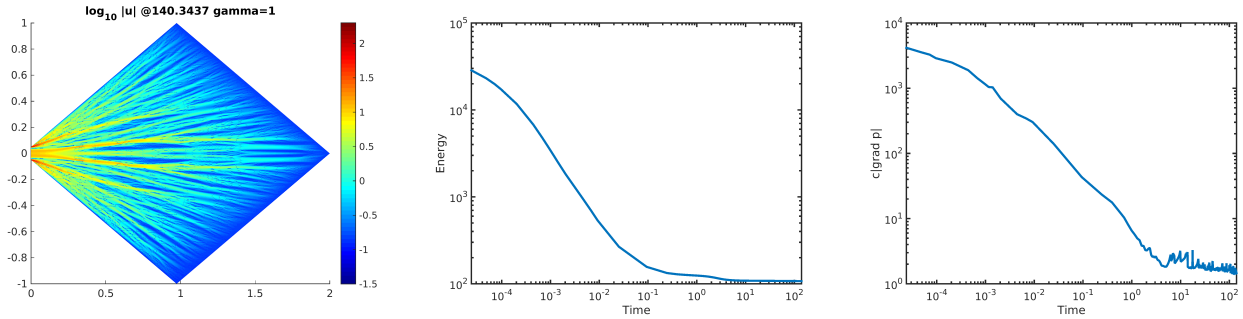

Figure 8: Velocity $|u_h^k|$ for $\gamma = 1$, $D = \frac{1}{1000}$ in a Log_{10}-scale, and corresponding evolution of the $\mathcal{E}_h(m_h^k)$ and $\|c|\nabla p|\|_{L^\infty(\Omega)}$ vs. time in a logarithmic scaling.

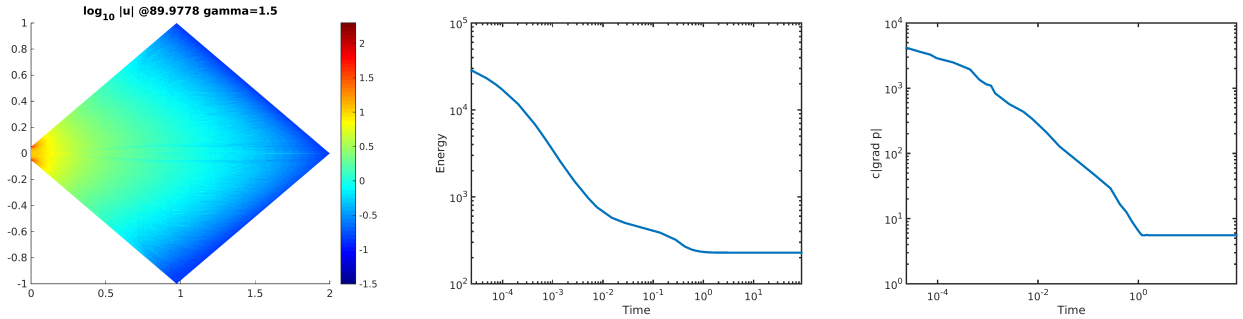

Figure 9: Velocity $|u_h^k|$ for $\gamma = \frac{3}{2}$, $D = \frac{1}{1000}$ in a Log_{10}-scale, and corresponding evolution of the $\mathcal{E}_h(m_h^k)$ and $\|c|\nabla p|\|_{L^\infty(\Omega)}$ vs. time in a logarithmic scaling.

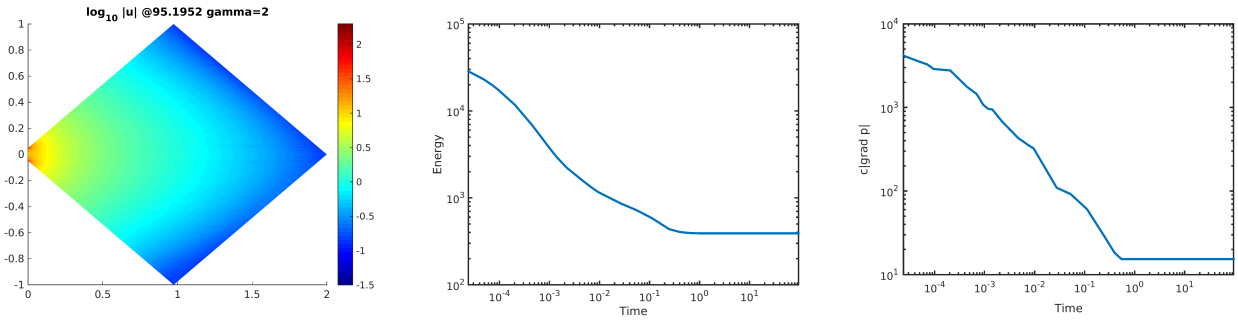

Figure 10: Velocity $|u_h^k|$ for $\gamma = 2$, $D = \frac{1}{1000}$ in a Log_{10}-scale, and corresponding evolution of the $\mathcal{E}_h(m_h^k)$ and $\|c|\nabla p|\|_{L^\infty(\Omega)}$ vs. time in a logarithmic scaling.

γ	$\frac{3}{5}$	$\frac{3}{4}$	1	$\frac{3}{2}$	2
t^k	121.5026	131.7063	140.3437	89.9778	95.1952
$\mathcal{E}_{h,t}^k$	-1.3×10^{-2}	-3.4×10^{-3}	-2.1×10^{-3}	-4.2×10^{-7}	-9.8×10^{-7}
$m_{h,t}^k$	2.1×10^{-1}	9.1×10^{-2}	1.1×10^{-1}	1.7×10^{-4}	1.1×10^{-5}
s_k	3.547262	4.153456	1.555159	1.181271	1.161117

Table 2: Stationarity and sparsity measures for different values of γ and $D = \frac{1}{1000}$, see Section 5.6. For $\gamma \in \{\frac{3}{2}, 2\}$ the change in the energy $\mathcal{E}_h$ is already within machine accuracy.

which is (4.19) for general values of c. For the minimization we use a gradient descent method with step-sizes chosen by the Armijo rule [12]. The iteration is stopped as soon as two subsequent iterates of the pressure, say p^k and p^{k+1}, satisfy $\|p^k - p^{k+1}\|_{H^1(\Omega)}/\|p^k\|_{H^1(\Omega)} < 10^{-15}$, i.e. they coincide up to round-off errors. Since the derivative of $\mathcal{F}_\alpha$ is discontinuous, one should in general use more general methods from convex optimization to ensure convergence of the minimization scheme, for instance proximal point methods [14]. However, in our example also the gradient descent method converged. We set $\gamma = 1/2$, $r = 1$ and $c = 50$. Moreover, we let $\mathcal{A} = \{x \in \mathbb{R}^2 : (x_1 - 1)^2 - x_2^2 < 1/4\} \subset \Omega$, see Figure 11. The stationary conductance m_0 is computed via (4.17) and satisfies

$$\|c\nabla p_0 \otimes c\nabla p_0 m_0 - |m_0|^{2(\gamma-1)} m_0\|_{L^2(\Omega)} \approx 2 \times 10^{-16},$$

which shows stationarity of the resulting solution (p_0, m_0) up to machine precision. The resulting stationary pressure and conductances are depicted in Figure 11.

In order to investigate the stability of the stationary state, we let η denote uniformly distributed random noise on $[-\frac{1}{2}, \frac{1}{2}]$, which we normalize such that $\|\eta\|_{L^2(\Omega)} = 1$. We set

$$m_\eta^0 = (1 + \frac{\eta}{1000}) m_0$$

as initial datum for our time-stepping scheme. The resulting evolution is depicted in Figure 12. Since the added noise is very small, the first picture in Figure 12 is visually identical to the absolute value $|m_0|$ of the unperturbed stationary solution. We observe that $m_\eta(t)$ does not converge to m_0, showing instability of the stationary state (p_0, m_0).

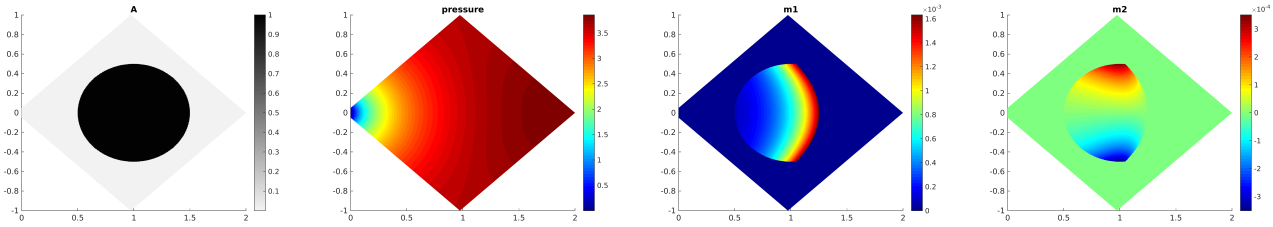

Figure 11: Stationary solution for $D = 0$ and $\gamma = 1/2$ via minimization of $\mathcal{F}_\alpha$ defined in (4.18). From left to right: Indicator function $\chi_{\mathcal{A}}$ of the set $\mathcal{A}$. Stationary pressure p_0. Stationary conductances $m_{0,1}$ and $m_{0,2}$.

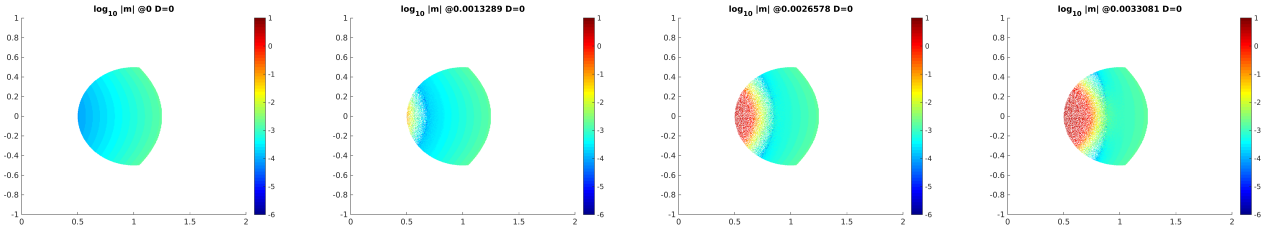

Figure 12: Evolution of the absolute value of the perturbed stationary solution m^0_η computed in Section 5.7 for different times in a logarithmic scale.

5.8 Finite time break-down for $\gamma < 1/2$

In Section 3.1 we have proven that m decays exponentially to zero for $-1 \leq \gamma \leq 1$, and that m becomes zero after a finite time if $\gamma < 1/2$ and the quantity cS is sufficiently small. In this case, the relaxation term $|m|^{2(\gamma-1)}m$ develops a singularity and is meaningless in the limit $|m| \to 0$. In the following we demonstrate numerically that this happens also in two dimensions, which complements the one-dimensional analysis of Section 3.1. In view of Section 3.1, we replace the homogeneous Dirichlet conditions for m by homogeneous Neumann conditions. For our test we set $c = 1$ and $\rho = 0$. To start with a strictly positive initial datum, we modify m^0 as follows

$$\tilde{m}^0_1 = m^0_1 + 10^{-3}, \quad \tilde{m}^0_2 = m^0_2,$$

and we define the extinction time as

$$T_{ex,\gamma} = \min\{t = t^k : \min_{x\in\Omega} |m^k_h(x)| < 10^{-8}\}.$$

The results are depicted in Table 3. We observe that the smaller γ the shorter the extinction time is. Monotone increase of $\gamma \mapsto T_{ex,\gamma}$ might be expected since for smaller values of γ the relaxation term becomes more singular as $m \to 0$. Therefore, for smaller γ the relaxation term dominates the activation term already for smaller times. In particular, $f_{\gamma,c}(m^k_h, \nabla p^k_h)$ acts like a sink. The threshold 10^{-8} seems to be somewhat arbitrary in the first place. However, in all our numerical simulations $\min_{x\in\Omega} |m^k_h(x)|$ decreased in a continuous fashion to values being approximately 10^{-6}, and then dropped below the threshold in one step. Therefore, all threshold values in the interval $[10^{-8}, 10^{-6}]$ would yield the same $T_{ex,\gamma}$ in these examples. This result indicates that Lemma 4 might be extended to multiple dimensions.

γ	0	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{2}{5}$
$T_{ex,\gamma}$	5.3×10^{-7}	2.6×10^{-6}	2.1×10^{-5}	2.1×10^{-4}
$\min_{x\in\Omega} \lvert m_h^k(x)\rvert$	5.9×10^{-9}	5.4×10^{-9}	3.7×10^{-12}	2.8×10^{-10}

Table 3: Extinction times $T_{ex,\gamma}$ for different values of γ.

Acknowledgment. BP is (partially) funded by the french "ANR blanche" project Kibord: ANR-13-BS01-0004" and by Institut Universitaire de France. PM acknowledges support of the Fondation Sciences Mathematiques de Paris in form of his Excellence Chair 2011. MS acknowledges support by ERC via Grant EU FP 7 - ERC Consolidator Grant 615216 LifeInverse.

References

[1] G. Albi, M. Artina, M. Fornasier and P. Markowich: *Biological transportation networks: modeling and simulation.* Preprint (2015).

[2] J. Aubin, and A. Cellina: *Differential Inclusions. Set Valued Maps and Viability Theory.* Springer-Verlag Berlin Heidelberg New York Tokyo, 1984.

[3] J. Àvila and A. Ponce: *Variants of Kato's inequality and removable singularities.* Journal d'Analyse Mathématique 91 (2003), pp. 143–178.

[4] D. Boffi, F. Brezzi, L. F. Demkowicz, R. G. Durn, R. S. Falk, and M. Fortin: *Mixed Finite Elements, Compatibility Conditions, and Applications* Springer, Berlin Heidelberg, 2008.

[5] L. Caffarelli, and N. Riviere, *The Lipschitz character of the stress tensor, when twisting an elastic plastic bar.* Arch. Rational Mech. Anal. 69 (1979), no. 1, pp. 3136.

[6] L. C. Evans: *Partial Differential Equations.* American Mathematical Society, Providence, Rhode Island, 1998.

[7] J. Haskovec, P. Markowich and B. Perthame: *Mathematical Analysis of a PDE System for Biological Network Formation.* Comm. PDE 40:5, pp. 918-956, 2015.

[8] D. Hu: *Optimization, Adaptation, and Initialization of Biological Transport Networks.* Notes from lecture (2013).

[9] D. Hu, private correspondence (2014).

[10] D. Hu and D. Cai: *Adaptation and Optimization of Biological Transport Networks.* Phys. Rev. Lett. 111 (2013), 138701.

[11] T. Koto: *IMEX Runge-Kutta schemes for reaction-diffusion equations.* Journal of Computational and Applied Mathematics (2008), Vol. 215, Issue 1, pp. 182–195.

[12] J. Nocedal, and S. J. Wright: *Numerical Optimization.* Springer, New York, 1999.

[13] R. Phelps: *Lectures on Maximal Monotone Operators.* Extracta Mathematicae 12 (1997), pp. 193-230.

[14] R. T. Rockafellar: *Monotone Operators and the Proximal Point Algorithm* SIAM J. Control and Optimization (1976), Vol. 14, No. 5, pp. 877–898.

[15] S. J. Ruuth: *Implicit-explicit methods for reaction-diffusion problems in pattern formation.* J. Math. Biol. (1995), 34:148–176.

[16] M. Safdari: *The regularity of some vector-valued variational inequalities with gradient constraints.* `arXiv:1501.05339`.

[17] M. Safdari: *The free boundary of variational inequalities with gradient constraints.* `arXiv:1501.05337`.